GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 24, 2004, 06:40:16 ; Search time 523 Seconds
(without alignments)
2371.843 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10723061/runat_21072004_090400_7098/app_query.fasta_1.455
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10723061_@CGN_1_1_885_@runat_21072004_090400_7098 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1486	100.0	1116	7	ADA70516	Ada70516 Rice gene
	2	1238	83.3	1083	6	ABZ12736	Abz12736 Arabidops
	3	1238	83.3	1083	7	ADA68129	Ada68129 Arabidops
	4	1238	83.3	1282	3	AAC42303	Aac42303 Arabidops
	5	1035	69.7	1086	3	AAC34458	Aac34458 Arabidops
	6	1035	69.7	1182	3	AAC39538	Aac39538 Arabidops
	7	1017	68.4	1688	3	AAC45046	Aac45046 Arabidops
	8	1012	68.1	1327	3	AAC36669	Aac36669 Arabidops
	9	949	63.9	1269	7	ADA71098	Ada71098 Rice gene
	10	763.5	51.4	110000	6	ABQ69245_13	Continuation (14 o
c	11	763.5	51.4	110000	6	ABQ67195_3	Continuation (4 of
	12	762	51.3	3003	4	ABL07155	Abl07155 Drosophil
	13	755.5	50.8	110000	6	ABA03041_13	Continuation (14 o
	14	751.5	50.6	5635	6	ABQ71010	Abq71010 Listeria
	15	743.5	50.0	858	6	ABN66178	Abn66178 Streptoco
	16	739.5	49.8	2000	7	ACC60718	Acc60718 Gene sequ
	17	739.5	49.8	2841	7	ABZ70970	Abz70970 Saccharom
	18	735.5	49.5	4474	9	ADB76850	Adb76850 A. gossyp
	19	735	49.5	3112	5	AAC91213	Aac91213 Human MTH
	20	735	49.5	3112	6	ABL68500	Abl68500 Kidney ca
	21	735	49.5	3112	6	ABN95605	Abn95605 Gene #210
c	22	735	49.5	3179	5	AAS66896	Aas66896 DNA encod
	23	731	49.2	3332	9	ADB57908	Adb57908 Toxicity-
	24	724	48.7	855	7	ABX06500	Abx06500 S. pneumo
	25	724	48.7	858	3	AAA05466	Aaa05466 Streptoco
	26	724	48.7	13440	2	AAV52261	Aav52261 Streptoco
	27	724	48.7	110000	7	ABS56454_06	Continuation (7 of
	28	721	48.5	852	6	ABN66177	Abn66177 Streptoco
	29	721	48.5	110000	6	ABN71527_04	Continuation (5 of
	30	713.5	48.0	2000	7	ACC60721	Acc60721 Gene sequ
	31	712.5	47.9	855	9	ADC94007	Adc94007 E. faeciu
	32	708.5	47.7	110000	6	ABA90521_08	Continuation (9 of
	33	704.5	47.4	7528	2	AAX12992	Aax12992 Enterococ
	34	704.5	47.4	7528	6	ABS98787	Abs98787 Enterococ
	35	687.5	46.3	855	6	ABK73707	Abk73707 Bacillus
	36	687	46.2	110000	3	AAA81489_7	Continuation (8 of
	37	687	46.2	172325	3	AAF21613	Aaf21613 Neisseria
c	38	681.5	45.9	110000	2	AAT42063_06	Continuation (7 of
	39	679.5	45.7	873	6	ABN91101	Abn91101 Staphyloc
	40	678	45.6	786	7	ABZ37976	Abz37976 N. gonorr
	41	675	45.4	876	7	ACF71018	Acf71018 Photorhab
	42	675	45.4	110000	7	ACF67367_41	Continuation (42 o
c	43	675	45.4	110000	7	ACF65388_06	Continuation (7 of
	44	670.5	45.1	882	7	ACF74867	Acf74867 Staphyloc
	45	670.5	45.1	9062	2	AAV74396	Aav74396 Staphyloc

ALIGNMENTS

RESULT 1
ADA70516
ID ADA70516 standard; DNA; 1116 BP.
XX
AC ADA70516;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 3839.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to

PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 3839; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 1116 BP; 268 A; 290 C; 296 G; 262 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.1e-154	Length:	1116
Score:	1486.00	Matches:	292
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-10-723-061-18 (1-292) x ADA70516 (1-1116)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 GCCAAGATCATTGATGGGAAGTTGGTGGCAAAGCAGATAAGAGAGGAAATCGCTGTTGAG 297

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 ATCGCCAAGATGAAGGATGCAATTGGGGTTGTGCCTGGGCTGGCAGTCATCCTAGTTGGG 357

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 TCAAGGAAGGATTCTCAAACGTATGTGCGCAACAAGAAGAAGGCATGCGAAGCGGTTGGT 417

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        418 ATCAAGTCATATGAGGTTAATTTGCCGGAAGACAGCTCTGAGGATGAGGTTCTCAAGCAC 477

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        478 ATCGCAACATTTAACAGTGATCCGTCGGTGCATGGCATCTTGGTTCAGTTGCCCCTACCT 537

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        538 CATCATATGAATGATGAGAACATTTTGAATGCTGTTAGTATTGAGAAGGATGTTGATGGA 597

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 TTTCATCCACTGAACATTGGACGACTTGCAATGCAAGGTCGGGATCCGTTCTTTGTTCCA 657

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 TGCACCCCTAAAGGATGCATGGAATTACTACACAGATATGGAGTTGAAATCAAAGGGAAG 717

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 AGAGCTGTTGTAATTGGGCGGAGCAATATTGTGGGGATGCCTGCTGCATTATTACTGCAA 777

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 AAAGCCAACGCAACTGTTAGCATTGTACATTCAAATACCAAGAAGCCTGAGGAAATAACA 837

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 AGACAAGCAGATATTGTTATCGCAGCTGTTGGAGTTGCTAATCTGGTCAGAGGGAGTTGG 897

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 ATAAAGCCTGGTGCTGCTATTATTGATGTTGGCATCAATCCAGTTGATGATCCAGAAAGC 957

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 CCTCGAGGTTATCGGCTGGTTGGAGATGTGTGCTACGAGGAGGCCTCCAAGATTGCAGGA 1017

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 CTAATCACGCCAGTTCCTGGTGGTGTCGGGCCAATGACAATTGCGATGCTTTTGTCGAAC 1077

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||||||||||||||||||||||||||||||||
Db       1078 ACACTTGAGTCAGCTAAAAGGATACACAAATTCAAA 1113
```

RESULT 2
ABZ12736
ID ABZ12736 standard; DNA; 1083 BP.
XX
AC ABZ12736;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 541.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US026685.
XX
PR 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.
XX
PS Claim 144; SEQ ID NO 541; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 1083 BP; 305 A; 230 C; 259 G; 289 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.22e-127	Length:	1083
Score:	1238.00	Matches:	231
Percent Similarity:	91.44%	Conservative:	36
Best Local Similarity:	79.11%	Mismatches:	25
Query Match:	83.31%	Indels:	0
DB:	6	Gaps:	0

US-10-723-061-18 (1-292) x ABZ12736 (1-1083)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::||||||||||||   |||||||||:::||||||:::||||||   :::|||
Db        205 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 264

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::||||||:::::::||||||||:::||||||||||||||||||||||||||
Db        265 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 324

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||   |||||||||||||||||||||||||||||:::::||||||
Db        325 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 384

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::||||||   |||   ||||||||||||||:::||||||||||||
Db        385 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 444

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::   ||||||   ||||||||||||||||||||||||||||||||||||||||
Db        445 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 504

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              ||||||:::::::::|||||||||||||||||||||||||||||||||||||||||||||
Db        505 TCGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATAGAGAAAGATGTTGACGGA 564

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||||||||||||||||||:::||||||:::|||    |||||||||
Db        565 TTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGAGAACCCTTATTCGTTCCT 624

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||:::||||||||||||||||||   :::||||||||||||||||
Db        625 TGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAACATTGAAATCAAAGGAAAG 684

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCAGCTGCTCTTTTACTGCAG 744

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::||||||||||||||||:::||||||   ||||||   ||||||||||||||||
Db        745 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 804

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::||||||||:::|||:::||||||||      |||:::|||||||||||||||
Db        805 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 864

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||   :::||||||||||||||||||||||:::||||||   :::
Db        865 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGAGGATCCAAGTGCT 924

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              |||||||||||||||||||||||:::|||||||||||||||||||||:::|||
Db        925 GCGCGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCA 984

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 GCCATCACACCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1044

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||   ||||||||||||||||||   |||:::
Db       1045 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1080
```

RESULT 3
ADA68129
ID ADA68129 standard; DNA; 1083 BP.
XX
AC ADA68129;
XX
DT 20-NOV-2003 (first entry)
XX
DE Arabidopsis thaliana gene, SEQ ID 393.
XX
KW Plant; bacterial infection; fungal infection; viral infection; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 393; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 1083 BP; 305 A; 230 C; 259 G; 289 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.22e-127	Length:	1083
Score:	1238.00	Matches:	231
Percent Similarity:	91.44%	Conservative:	36
Best Local Similarity:	79.11%	Mismatches:	25
Query Match:	83.31%	Indels:	0
DB:	7	Gaps:	0

US-10-723-061-18 (1-292) x ADA68129 (1-1083)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::||||||||||||   |||||||||:::||||||:::||||||   :::|||
Db        205 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 264

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::||||||:::::::||||||||:::||||||||||||||||||||||||||
Db        265 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 324

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||   |||||||||||||||||||||||||||||:::::||||||
Db        325 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 384

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::||||||   |||   ||||||||||||||:::||||||||||||
Db        385 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 444

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::   ||||||   ||||||||||||||||||||||||||||||||||||||||
Db        445 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 504

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              ||||||:::::::::|||||||||||||||||||||||||||||||||||||||||||||
Db        505 TCGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATAGAGAAAGATGTTGACGGA 564
```

```
Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||||||||||||||:::|||||||:::|||   |||||||||
Db        565 TTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGAGAACCCTTATTCGTTCCT 624

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||:::||||||||||||||||||   :::||||||||||||||||
Db        625 TGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAACATTGAAATCAAAGGAAAG 684

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCAGCTGCTCTTTTACTGCAG 744

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::|||||||||||||||:::||||||   ||||||   ||||||||||||||||
Db        745 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 804

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||:::|||:::|||||||||      |||:::|||||||||||||||
Db        805 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 864

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||   :::|||||||||||||||||||||||||:::|||||||   :::
Db        865 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGAGGATCCAAGTGCT 924

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                  |||||||||||||||||||||:::|||||||||||||||||||||:::|||
Db        925 GCGCGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCA 984

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 GCCATCACACCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1044

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||   ||||||||||||||||||   |||:::
Db       1045 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1080

RESULT 4
AAC42303
ID   AAC42303 standard; DNA; 1282 BP.
XX
AC   AAC42303;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 35038.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic pathway;
KW   promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   14-MAY-1999;   99US-0134370P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137528P.
PR   04-JUN-1999;   99US-0137502P.
PR   07-JUN-1999;   99US-0137724P.
PR   08-JUN-1999;   99US-0138094P.
PR   10-JUN-1999;   99US-0138540P.
PR   10-JUN-1999;   99US-0138847P.
PR   14-JUN-1999;   99US-0139119P.
PR   16-JUN-1999;   99US-0139452P.
PR   16-JUN-1999;   99US-0139453P.
PR   17-JUN-1999;   99US-0139492P.
PR   18-JUN-1999;   99US-0139454P.
PR   18-JUN-1999;   99US-0139455P.
PR   18-JUN-1999;   99US-0139456P.
PR   18-JUN-1999;   99US-0139457P.
PR   18-JUN-1999;   99US-0139458P.
PR   18-JUN-1999;   99US-0139459P.
PR   18-JUN-1999;   99US-0139460P.
PR   18-JUN-1999;   99US-0139461P.
PR   18-JUN-1999;   99US-0139462P.
PR   18-JUN-1999;   99US-0139463P.
PR   18-JUN-1999;   99US-0139750P.
PR   18-JUN-1999;   99US-0139763P.
PR   21-JUN-1999;   99US-0139817P.
PR   22-JUN-1999;   99US-0139899P.
PR   23-JUN-1999;   99US-0140353P.
PR   23-JUN-1999;   99US-0140354P.
PR   24-JUN-1999;   99US-0140695P.
PR   28-JUN-1999;   99US-0140823P.
PR   29-JUN-1999;   99US-0140991P.
PR   30-JUN-1999;   99US-0141287P.
PR   01-JUL-1999;   99US-0141842P.
PR   01-JUL-1999;   99US-0142154P.
PR   02-JUL-1999;   99US-0142055P.
PR   06-JUL-1999;   99US-0142390P.
PR   08-JUL-1999;   99US-0142803P.
PR   09-JUL-1999;   99US-0142920P.
PR   12-JUL-1999;   99US-0142977P.
PR   13-JUL-1999;   99US-0143542P.
PR   14-JUL-1999;   99US-0143624P.
PR   15-JUL-1999;   99US-0144005P.
PR   16-JUL-1999;   99US-0144085P.
PR   16-JUL-1999;   99US-0144086P.
PR   19-JUL-1999;   99US-0144325P.
PR   19-JUL-1999;   99US-0144331P.
PR   19-JUL-1999;   99US-0144332P.
PR   19-JUL-1999;   99US-0144333P.
PR   19-JUL-1999;   99US-0144334P.
PR   19-JUL-1999;   99US-0144335P.
PR   20-JUL-1999;   99US-0144352P.
PR   20-JUL-1999;   99US-0144632P.
PR   20-JUL-1999;   99US-0144884P.
PR   21-JUL-1999;   99US-0144814P.
```

PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.

PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.:	4.08e-127	Length:	1282
Score:	1238.00	Matches:	231
Percent Similarity:	91.44%	Conservative:	36
Best Local Similarity:	79.11%	Mismatches:	25
Query Match:	83.31%	Indels:	0
DB:	3	Gaps:	0

US-10-723-061-18 (1-292) x AAC42303 (1-1282)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::|||||||||||||    |||||||||:::||||||:::||||||   :::|||
Db        404 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 463

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::||||||::::::||||||||||:::||||||||||||||||||||||||||
Db        464 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 523

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||   ||||||||||||||||||||||||||||||:::::||||||
Db        524 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 583

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::||||||    |||    ||||||||||||||||:::||||||||||||
Db        584 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 643

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::  ||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        644 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 703

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              ||||||::::::::::||||||||||||||||||||||||||||||||||||||||||||
Db        704 TCGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATAGAGAAAGATGTTGACGGA 763

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||||||||||||||||||:::|||||:::|||   |||||||||
Db        764 TTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGAGAACCCTTATTCGTTCCT 823

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||||||||||||||||:::|||||||||||||||||||   :::|||||||||||||||
Db        824 TGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAACATTGAAATCAAAGGAAAG 883

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        884 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCAGCTGCTCTTTTACTGCAG 943

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::|||||||||||||||:::||||||   ||||||   ||||||||||||||||
Db        944 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 1003

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::||||||||:::|||:::||||||||      |||:::|||||||||||||||
Db       1004 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 1063
```

```
Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||   :::|||||||||||||||||||||||:::|||||||   :::
Db       1064 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGAGGATCCAAGTGCT 1123

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||:::|||||||||||||||||||||:::|||
Db       1124 GCGCGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCA 1183

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1184 GCCATCACACCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1243

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||   ||||||||||||||||||   |||:::
Db       1244 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1279

RESULT 5
AAC34458
ID   AAC34458 standard; DNA; 1086 BP.
XX
AC   AAC34458;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 6720.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic pathway;
KW   promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   14-MAY-1999;   99US-0134370P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137528P.
PR   04-JUN-1999;   99US-0137502P.
PR   07-JUN-1999;   99US-0137724P.
PR   08-JUN-1999;   99US-0138094P.
PR   10-JUN-1999;   99US-0138540P.
PR   10-JUN-1999;   99US-0138847P.
PR   14-JUN-1999;   99US-0139119P.
PR   16-JUN-1999;   99US-0139452P.
PR   16-JUN-1999;   99US-0139453P.
PR   17-JUN-1999;   99US-0139492P.
PR   18-JUN-1999;   99US-0139454P.
PR   18-JUN-1999;   99US-0139455P.
PR   18-JUN-1999;   99US-0139456P.
PR   18-JUN-1999;   99US-0139457P.
PR   18-JUN-1999;   99US-0139458P.
PR   18-JUN-1999;   99US-0139459P.
PR   18-JUN-1999;   99US-0139460P.
PR   18-JUN-1999;   99US-0139461P.
PR   18-JUN-1999;   99US-0139462P.
PR   18-JUN-1999;   99US-0139463P.
PR   18-JUN-1999;   99US-0139750P.
PR   18-JUN-1999;   99US-0139763P.
PR   21-JUN-1999;   99US-0139817P.
PR   22-JUN-1999;   99US-0139899P.
PR   23-JUN-1999;   99US-0140353P.
PR   23-JUN-1999;   99US-0140354P.
PR   24-JUN-1999;   99US-0140695P.
PR   28-JUN-1999;   99US-0140823P.
PR   29-JUN-1999;   99US-0140991P.
PR   30-JUN-1999;   99US-0141287P.
PR   01-JUL-1999;   99US-0141842P.
PR   01-JUL-1999;   99US-0142154P.
PR   02-JUL-1999;   99US-0142055P.
PR   06-JUL-1999;   99US-0142390P.
PR   08-JUL-1999;   99US-0142803P.
PR   09-JUL-1999;   99US-0142920P.
PR   12-JUL-1999;   99US-0142977P.
PR   13-JUL-1999;   99US-0143542P.
PR   14-JUL-1999;   99US-0143624P.
PR   15-JUL-1999;   99US-0144005P.
PR   16-JUL-1999;   99US-0144085P.
PR   16-JUL-1999;   99US-0144086P.
PR   19-JUL-1999;   99US-0144325P.
PR   19-JUL-1999;   99US-0144331P.
PR   19-JUL-1999;   99US-0144332P.
PR   19-JUL-1999;   99US-0144333P.
PR   19-JUL-1999;   99US-0144334P.
PR   19-JUL-1999;   99US-0144335P.
PR   20-JUL-1999;   99US-0144352P.
PR   20-JUL-1999;   99US-0144632P.
PR   20-JUL-1999;   99US-0144884P.
PR   21-JUL-1999;   99US-0144814P.
PR   21-JUL-1999;   99US-0145086P.
PR   21-JUL-1999;   99US-0145088P.
PR   22-JUL-1999;   99US-0145085P.
PR   22-JUL-1999;   99US-0145087P.
PR   22-JUL-1999;   99US-0145089P.
PR   22-JUL-1999;   99US-0145192P.
PR   23-JUL-1999;   99US-0145145P.
PR   23-JUL-1999;   99US-0145218P.
PR   23-JUL-1999;   99US-0145224P.
PR   26-JUL-1999;   99US-0145276P.
PR   27-JUL-1999;   99US-0145913P.
PR   27-JUL-1999;   99US-0145918P.
PR   27-JUL-1999;   99US-0145919P.
PR   28-JUL-1999;   99US-0145951P.
PR   02-AUG-1999;   99US-0146386P.
PR   02-AUG-1999;   99US-0146388P.
PR   02-AUG-1999;   99US-0146389P.
PR   03-AUG-1999;   99US-0147038P.
PR   04-AUG-1999;   99US-0147204P.
```

PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.:	1.02e-104	Length:	1086
Score:	1035.00	Matches:	194
Percent Similarity:	82.99%	Conservative:	45
Best Local Similarity:	67.36%	Mismatches:	49
Query Match:	69.65%	Indels:	0
DB:	3	Gaps:	0

US-10-723-061-18 (1-292) x AAC34458 (1-1086)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||||||||||||||   :::|||      ||||||   |||||||||   |||
Db         67 GCGAAGATAATCGACGGCAAGGCGATTGCTCATACCATCAGATCAGAGATCGCCGAGGAA 126

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::      :::   :::      |||   ||||||||||||||||:::::||||||
Db        127 GTTCGCGGTCTATCTGAGAAACACGGCAAGGTCCCAGGACTAGCTGTAGTTATTGTTGGG 186

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||      |||:::|||||||||      ||||||
Db        187 AGCCGAAAGGATTCACAGACCTATGTGAATACGAAGAGGAAAGCGTGCGCTGAGGTTGGG 246

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||::::::|||   |||||||||:::   ||||||   :::::::::
Db        247 ATTAAGTCATTTGACGTGGGCCTACCAGAGGAAGTTTCTGAAGCTGATCTTATTAGCAAA 306

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::      ||||||:::|||   ||||||||||||||||||||||||||||||||||||
Db        307 GTTCATGAACTAAATTCAAATCCGGATGTCCATGGCATATTAGTTCAACTCCCATTGCCG 366

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||:::|||:::|||||||||||||   |||:::||||||:::|||||||||||||||
Db        367 AAACATATTAATGAGGAGAATATATTGGGTGCAATCAGCATTGATAAAGATGTCGATGGC 426

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||||||:::|||||||||:::||||||:::|||   |||:::|||
Db        427 TTCCATCCTTTGAATATTGGTAAGCTAGCCATGAAAGGCAGAGAACCCCTCTTCCTTCCT 486

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||||||||||||||||:::|||||||||   |||   ||||||:::||||||||:::
Db        487 TGCACCCCAAAGGGATGTTTGGAACTCCTAGCTAGAAGCGGCGTAAAGATAAAGGGGCAA 546

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlyGln 180
              ||||||||||||:::|||||||||||||||||||||:::|||   :::||||||||||
Db        547 CGAGCAGTTGTTGTAGGTCGGAGTAACATTGTTGGTTTGCCCGTTTCACTTCTTTTGCTC 606

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::||||||||:::   |||||||||:::||||||   ||||||   |||
Db        607 AAGGCTGATGCTACTGTCACAACTGTACATTCTCACACCAAGGATCCTGAGGCTATCATA 666

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||||   |||   |||:::::::::::|||:::|||
Db        667 CGGGAAGCTGACATTGTTATTGCTGCATGCGGACAAGCCCACATGATTAAGGGCAACTGG 726

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||:::|||||||||||   |||   |||   ||||||
Db        727 ATAAAGCCAGGGGCTGCAGTAATTGATGTTGGAACTAATGCAGTCAGCGACCCGAGCAAG 786

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                  ||||||||||||||||||||||||   :::   ||||||||||||:::||||||
Db        787 AAATCAGGATACCGGTTGGTTGGAGATGTTGATTTCGCAGAAGCTTCAAAAGTTGCAGGT 846

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||:::||||||||||||   |||
Db        847 TTCATAACTCCGGTCCCTGGTGGTGTAGGCCCAATGACAGTGGCAATGCTTCTCAGGAAC 906

Qy        281 ThrLeuGluSerAlaLysArgIle 288
              |||::::::   |||||||||:::
Db        907 ACCGTAGACGGTGCCAAGCGTGTC 930
```

RESULT 6
AAC39538

ID AAC39538 standard; DNA; 1182 BP.
XX
AC AAC39538;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 25000.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.

PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.:	1.15e-104	Length:	1182
Score:	1035.00	Matches:	194
Percent Similarity:	82.99%	Conservative:	45
Best Local Similarity:	67.36%	Mismatches:	49
Query Match:	69.65%	Indels:	0
DB:	3	Gaps:	0

US-10-723-061-18 (1-292) x AAC39538 (1-1182)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              ||||||||||||||||||||   :::|||     ||||||   |||||||||    |||
Db        114 GCGAAGATAATCGACGGCAAGGCGATTGCTCATACCATCAGATCAGAGATCGCCGAGGAA 173

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::    :::   :::      |||   ||||||||||||||||||:::::||||||
Db        174 GTTCGCGGTCTATCTGAGAAACACGGCAAGGTCCCAGGACTAGCTGTAGTTATTGTTGGG 233

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||      |||:::|||||||||      ||||||
Db        234 AGCCGAAAGGATTCACAGACCTATGTGAATACGAAGAGGAAAGCGTGCGCTGAGGTTGGG 293

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::::|||   |||||||||:::   ||||||   :::::::::
Db        294 ATTAAGTCATTTGACGTGGGCCTACCAGAGGAAGTTTCTGAAGCTGATCTTATTAGCAAA 353

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::          ||||||:::|||   ||||||||||||||||||||||||||||||||
Db        354 GTTCATGAACTAAATTCAAATCCGGATGTCCATGGCATATTAGTTCAACTCCCATTGCCG 413

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||:::|||:::||||||||||||   |||:::||||||:::|||||||||||||||
Db        414 AAACATATTAATGAGGAGAATATATTGGGTGCAATCAGCATTGATAAAGATGTCGATGGC 473

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||:::|||||||:::|||||:::|||   |||:::|||
Db        474 TTCCATCCTTTGAATATTGGTAAGCTAGCCATGAAAGGCAGAGAACCCCTCTTCCTTCCT 533

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||:::|||||||||   |||   ||||||:::||||||||||:::
Db        534 TGCACCCCAAAGGGATGTTTGGAACTCCTAGCTAGAAGCGGCGTAAAGATAAAGGGGCAA 593

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||:::|||||||||||||||||||:::|||   :::|||||||||
Db        594 CGAGCAGTTGTTGTAGGTCGGAGTAACATTGTTGGTTTGCCCGTTTCACTTCTTTTGCTC 653

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::|||||||||:::   |||||||||:::||||||   ||||||   |||
Db        654 AAGGCTGATGCTACTGTCACAACTGTACATTCTCACACCAAGGATCCTGAGGCTATCATA 713

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||||   |||   |||:::::::::::|||:::|||
Db        714 CGGGAAGCTGACATTGTTATTGCTGCATGCGGACAAGCCCACATGATTAAGGGCAACTGG 773

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||:::|||||||||||   |||   |||   ||||||
Db        774 ATAAAGCCAGGGGCTGCAGTAATTGATGTTGGAACTAATGCAGTCAGCGACCCGAGCAAG 833

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                  ||||||||||||||||||||||||   :::   ||||||||||||:::||||||
Db        834 AAATCAGGATACCGGTTGGTTGGAGATGTTGATTTCGCAGAAGCTTCAAAAGTTGCAGGT 893

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              |||||||||||||||||||||||||||||||||||:::||||||||||||   |||
Db        894 TTCATAACTCCGGTCCCTGGTGGTGTAGGCCCAATGACAGTGGCAATGCTTCTCAGGAAC 953

Qy        281 ThrLeuGluSerAlaLysArgIle 288
              |||::::::   |||||||||:::
Db        954 ACCGTAGACGGTGCCAAGCGTGTC 977
```

RESULT 7
AAC45046
ID AAC45046 standard; DNA; 1688 BP.
XX
AC AAC45046;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 45095.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.

PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.:	1.86e-102	Length:	1688
Score:	1017.00	Matches:	193
Percent Similarity:	80.07%	Conservative:	36
Best Local Similarity:	67.48%	Mismatches:	57
Query Match:	68.44%	Indels:	0
DB:	3	Gaps:	0

US-10-723-061-18 (1-292) x AAC45046 (1-1688)

```
Qy          3 IleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIleAla 22
              :::|||||||||   ::::::|||:::::::||||||   :::|||       |||:::
Db        631 GTGATAGATGGTAATGTTATCGCTGAGGAAATCAGAACAAAGATTATTAGTGAAGTGGGG 690

Qy         23 LysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySerArg 42
              |||||||||   |||:::|||   |||||||||||||||||||:::|||||||||   :::
Db        691 AAGATGAAAAAGGCTGTTGGAAAAGTCCCTGGTCTTGCTGTTGTTTTGGTTGGTGAACAA 750

Qy         43 LysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIleLys 62
              :::|||||||||||||||||||||||||||   ||||||||||||      |||||||||
Db        751 AGAGACTCGCAAACTTATGTTCGCAACAAAATTAAAGCTTGTGAAGAAACTGGCATTAAA 810

Qy         63 SerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAla 82
              |||               ||||||||||||   :::|||   :::::::::       :::
Db        811 TCTGTTTTGGCTGAATTACCAGAGGATTGTACTGAAGGACAGATTATCAGTGTATTGAGA 870

Qy         83 ThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHis 102
                 ||||||   |||   |||:::|||||||||||||||||||||||||||||||   |||
Db        871 AAATTCAATGAAGACACGTCTATTCACGGAATTCTTGTGCAACTTCCTTTACCCCAACAT 930

Qy        103 MetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHis 122
              :::|||:::   ||||||||   |||   :::|||||||||||||||||||||||||||
Db        931 CTTAATGAATCAAAAATCTTGAATATGGTGAGATTGGAGAAAGATGTTGATGGGTTTCAT 990

Qy        123 ProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThr 142
              |||||||||:::|||   |||||||||:::|||||:::|||   ||||||   ||||||
Db        991 CCATTAAACGTGGGGAATCTTGCAATGAGAGGGAGGGAACCGCTGTTTGTATCTTGCACT 1050

Qy        143 ProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAla 162
              ||||||||||:::|||||||||   |||   ||||||||||||   ||||||   |||
Db       1051 CCTAAGGGCTGTGTGGAGTTGTTGATAAGAACGGGTGTTGAAATAGCAGGAAAAAATGCT 1110

Qy        163 ValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLysAla 182
              |||||||||||||||||||||||||||:::|||   :::||||||||||||:::
Db       1111 GTGGTCATAGGCAGAAGCAACATCGTTGGACTACCAATGTCTTTATTATTGCAGAGGCAT 1170

Qy        183 AsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGln 202
              :::||||||||||||   ||||||:::   ||||||   ||||||   |||||||||:::
Db       1171 GACGCGACAGTAAGCACGGTGCATGCGTTTACAAAGGATCCAGAGCACATTACTAGGAAG 1230

Qy        203 AlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLys 222
              ||||||||||||||||||||   |||:::   ||||||||||||||||||||:::|||
Db       1231 GCGGATATCGTTATAGCAGCAGCGGGTATACCTAATTTAGTTCGTGGAAGTTGGCTTAAG 1290

Qy        223 ProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerProArg 242
              |||||||||   :::||||||||||||      ||||||:::|||
Db       1291 CCTGGAGCAGTTGTCATTGATGTTGGAACAACTCCAGTCGAGGATAGTAGCTGTGAGTTT 1350

Qy        243 GlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIle 262
              ||||||||||||||||||||||||||||||||||||||      :::|||      |||
Db       1351 GGTTACCGTCTTGTCGGGGATGTATGTTACGAGGAAGCGTTAGGTGTTGCCTCGGCTATC 1410

Qy        263 ThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeu 282
              |||||||||||||||||||||||||||||||||   |||||||||   |||||||||
Db       1411 ACTCCCGTACCGGGAGGAGTTGGACCTATGACGATCACCATGCTACTATGCAATACTTTG 1470

Qy        283 GluSerAlaLysArgIle 288
              |||:::|||||||||||||
Db       1471 GAGGCTGCTAAGCGGATC 1488
```

RESULT 8
AAC36669
ID AAC36669 standard; DNA; 1327 BP.
XX
AC AAC36669;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 14649.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.

```
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   25-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.

Alignment Scores:
Pred. No.:              4.76e-102       Length:         1327
Score:                  1012.00         Matches:        192
Percent Similarity:     80.07%          Conservative:   37
Best Local Similarity:  67.13%          Mismatches:     57
Query Match:            68.10%          Indels:         0
DB:                     3               Gaps:           0

US-10-723-061-18 (1-292) x AAC36669 (1-1327)

Qy          3 IleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIleAla 22
              :::|||||||||   ::::::|||::::::||||||   :::|||       |||:::
Db        270 GTGATAGATGGTAATGTTATCGCTGAGGAAATCAGAACAAAGATTATTAGTGAAGTGGGG 329

Qy         23 LysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySerArg 42
              |||||||||   |||:::|||   |||||||||||||||||:::|||||||||   :::
Db        330 AAGATGAAAAAGGCTGTTGGAAAAGTCCCTGGTCTTGCTGTTGTTTTGGTTGGTGAACAA 389

Qy         43 LysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIleLys 62
              :::|||||||||||||||||||||||||||||   ||||||||||||      |||||||||
Db        390 AGAGACTCGCAAACTTATGTTCGCAACAAAATTAAAGCTTGTGAAGAAACTGGCATTAAA 449

Qy         63 SerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAla 82
              |||             ||||||||||||   :::|||   :::::::::      :::
Db        450 TCTGTTTTGGCTGAATTACCAGAGGATTGTACTGAAGGACAGATTATCAGTGTATTGAGA 509

Qy         83 ThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHis 102
              ||||||   |||   |||:::|||||||||||||||||||||||||||||||   |||
Db        510 AAATTCAATGAAGACACGTCTATTCACGGAATTCTTGTGCAACTTCCTTTACCCCAACAT 569

Qy        103 MetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHis 122
              :::::::::      |||||||||   |||   :::|||||||||||||||||||||||
Db        570 CTTGATGAATCAAAAATCTTGAATATGGTGAGATTGGAGAAAGATGTTGATGGGTTTCAT 629

Qy        123 ProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThr 142
              |||||||||:::|||   |||||||||:::||||||:::|||   ||||||   ||||||
Db        630 CCATTAAACGTGGGGAATCTTGCAATGAGAGGGAGGGAACCGCTGTTTGTGTCTTGCACT 689

Qy        143 ProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAla 162
              ||||||||||:::|||||||||   |||   ||||||||||||   ||||||   |||
Db        690 CCTAAGGGCTGTGTGGAGTTGTTGATAAGAACGGGTGTTGAAATAGCAGGAAAAAATGCT 749

Qy        163 ValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLysAla 182
              |||||||||||||||||||||||||||:::|||   :::||||||||||||:::
Db        750 GTGGTCATAGGCAGAAGCAACATCGTTGGACTACCAATGTCTTTATTATTGCAGAGGCAT 809

Qy        183 AsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGln 202
              :::||||||||||||   ||||||:::   ||||||   ||||||   |||||||||:::
Db        810 GACGCGACAGTAAGCACGGTGCATGCGTTTACAAAGGATCCAGAGCACATTACTAGGAAG 869

Qy        203 AlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLys 222
              |||||||||||||||||||   |||:::   ||||||||||||||||||||:::|||
Db        870 GCGGATATCGTTATAGCAGCAGCGGGTATACCTAATTTAGTTCGTGGAAGTTGGCTTAAG 929

Qy        223 ProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerProArg 242
              |||||||||   :::||||||||||||      ||||||:::|||
Db        930 CCTGGAGCAGTTGTCATTGATGTTGGAACAACTCCAGTCGAGGATAGTAGCTGTGAGTTT 989

Qy        243 GlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIle 262
              ||||||||||||||||||||||||||||||||||||||||||||      :::|||      |||
Db        990 GGTTACCGTCTTGTCGGGGATGTATGTTACGAGGAAGCGTTAGGTGTTGCCTCGGCTATC 1049

Qy        263 ThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeu 282
              |||||||||||||||||||||||||||||||||||||   ||||||||   |||||||||
Db       1050 ACTCCCGTACCGGGAGGAGTTGGACCTATGACGATCACCATGCTACTATGCAATACTTTG 1109

Qy        283 GluSerAlaLysArgIle 288
              |||:::||||||||||||
Db       1110 GAGGCTGCTAAGCGGATC 1127

RESULT 9
ADA71098
ID   ADA71098 standard; DNA; 1269 BP.
XX
AC   ADA71098;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Rice gene, SEQ ID 4421.
XX
KW   Plant; bacterial infection; fungal infection; viral infection; rice;
KW   gene; ds.
XX
OS   Oryza sativa.
XX
PN   WO2003000898-A1.
XX
PD   03-JAN-2003.
XX
PF   22-JUN-2001; 2001WO-IB001105.
XX
PR   22-JUN-2001; 2001WO-IB001105.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Chang H,  Chen W,  Cooper B,  Glazebrook J,  Goff SA,  Hou Y;
PI   Katagiri F,  Quan S,  Tao Y,  Whitham S,  Xie Z,  Zhu T,  Zou G;
XX
DR   WPI; 2003-175290/17.
XX
PT   Identifying at least one gene involved in plant resistance or response to
PT   pathogenic infection for conferring resistance or tolerance to a plant to
PT   bacterial, fungal or viral infection by determining or detecting plant
PT   gene expression.
XX
PS   Claim 6; SEQ ID NO 4421; 899pp; English.
XX
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1
CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,
CC   or in a corresponding incompatible or compatible interaction. (M1) is
CC   useful for conferring resistance to resistance or tolerance to a plant to
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 1269 BP; 271 A; 305 C; 378 G; 309 T; 0 U; 6 Other;

Alignment Scores:
Pred. No.:              4.3e-95         Length:         1269
Score:                  949.00          Matches:        182
Percent Similarity:     78.67%          Conservative:   43
Best Local Similarity:  63.64%          Mismatches:     55
Query Match:            63.86%          Indels:         6
DB:                     7               Gaps:           1

US-10-723-061-18 (1-292) x ADA71098 (1-1269)
```

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   |||||||||||||||   ||||||:::   ||||||   :::||||||   |||
Db        391 GCGACCATTATTGATGGCAAGTCTGTCGCGGAAGACATAAGGTTTCAGATCGCTGAAGAA 450

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::   :::|||||||:::|||:::|||   |||||||||||||||||||:::||||||||||
Db        451 GTTCGTCAAATGAAAAATGCAGTGGGGCATGTGCCTGGCCTAGCTGTTGTATTGGTCGGC 510

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
                ||:::||||||:::::::|||||||||   |||   |||   ||||||   ||||||
Db        511 GATAGAAGGGACTCTGAGTCATACGTGCGATACAAAATTAAGGGCTGTGAGGAGGTTGGA 570

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||             ||||||   :::   :::||||||   |||:::
Db        571 ATAAAGTCCTTGTTGGCGGAGTTGCCTGGGAACTGTACGGAAGACGTAGTGGTGGATTCT 630

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::   ||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        631 GTGTCAAGATTCAATGAAGATCCATCTGTTCATGGCATTCTTGTGCAACTACCTCTACCA 690

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                                  |||:::|||:::|||:::||||||||||||||||||||
Db        691 CAGGAT-----------------TTGAGCGCTATTAGCCTAGAGAAGGATGTTGATGGT 732

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||:::|||   ||||||::::::   |||   |||   |||||||||
Db        733 TTCCATCCATTGAATGTTGGTAATCTTGCCCTCAGAAGCCGGAAGCCTTTATTTGTGCCC 792

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||       |||   |||:::|||||||||   :::   |||:::|||:::   ||||||
Db        793 TGTGCTGCAAAGGCTTGCCTTGAGTTATTGCTCCAGTCTGGGATTGAACTCATGGGGAAA 852

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
                  ||||||||||||||   :::||||||:::|||   :::||||||||||||
Db        853 CATGTCACTGTGATTGGGAGAAGCAAAGTTGTTGGGTTGCCCACTTCTTTACTTTTACAG 912

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::|||||||||||||||:::|||:::   |||       ||||||||||||||
Db        913 AGACACCATGCCACTGTTAGTATTATCCATGCTTTCACAACAAATCCAGAAGAGATTACC 972

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              ||||||:::|||||||||||||:::|||   ||||||||||||||||||||||||||||||
Db        973 CGTCAATCTGATATTGTGATCTCAGCTGCTGGAGTGGCCAATCTTGTAAGAGGAAGCTGG 1032

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              :::|||   ||||||   :::||||||||||||   ||||||::::::|||||   |||
Db       1033 TTGAAGAAAGGTGCAGTTGTGATCGATGTTGGAACAAACCCAATTGAGGATCCAACCAGT 1092

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                 ||||||||||||   ||||||||||||:::|||||||||   |||:::|||
Db       1093 GATTACGGCTATCGATTAACTGGAGATGTCTGCTTCGAAGAAGCTGTGAAGTTGGCCTCT 1152

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
                 ||||||||||||||||||||||||||||||:::||||||||||||||||||:::|||
Db       1153 GCTATAACTCCAGTTCCTGGTGGTGTGGGACCGGTGACCATTGCGATGCTTCTTGCGAAT 1212

Qy        281 ThrLeuGluSerAlaLys 286
              ||||||:::||||||||
Db       1213 ACACTTGACTCAGCGAAA 1230

RESULT 10
ABQ69245_13
Continuation (14 of 31) of ABQ69245 from base 1300001 (Listeria innocua DNA sequence #68
WP Sequence split into 31 fragments  LOCUS ABQ69245 Accession Abq69245
WP      Fragment Name       Begin     End
WP      ABQ69245_00             1     110000
WP      ABQ69245_01        100001     210000
WP      ABQ69245_02        200001     310000
WP      ABQ69245_03        300001     410000
WP      ABQ69245_04        400001     510000
WP      ABQ69245_05        500001     610000
WP      ABQ69245_06        600001     710000
WP      ABQ69245_07        700001     810000
WP      ABQ69245_08        800001     910000
WP      ABQ69245_09        900001    1010000
WP      ABQ69245_10       1000001    1110000
WP      ABQ69245_11       1100001    1210000
WP      ABQ69245_12       1200001    1310000
WP      ABQ69245_13       1300001    1410000
WP      ABQ69245_14       1400001    1510000
WP      ABQ69245_15       1500001    1610000
WP      ABQ69245_16       1600001    1710000
WP      ABQ69245_17       1700001    1810000
WP      ABQ69245_18       1800001    1910000
WP      ABQ69245_19       1900001    2010000
WP      ABQ69245_20       2000001    2110000
WP      ABQ69245_21       2100001    2210000
WP      ABQ69245_22       2200001    2310000
WP      ABQ69245_23       2300001    2410000
WP      ABQ69245_24       2400001    2510000
WP      ABQ69245_25       2500001    2610000
WP      ABQ69245_26       2600001    2710000
WP      ABQ69245_27       2700001    2810000
WP      ABQ69245_28       2800001    2910000
WP      ABQ69245_29       2900001    3010000
WP      ABQ69245_30       3000001    3011208

Alignment Scores:
Pred. No.:              7.84e-72        Length:         110000
Score:                  763.50          Matches:        151
Percent Similarity:     72.66%          Conservative:   59
Best Local Similarity:  52.25%          Mismatches:     70
Query Match:            51.38%          Indels:         9
DB:                     6               Gaps:           4

US-10-723-061-18 (1-292) x ABQ69245_13 (1-110000)

Qy          2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
              :::|||||||||||||||   :::||||||:::|||:::|||:::::::       |||:::
Db      90282 GAAATTATTGATGGCAAAAAGTTAGCCAAAGAAATCCAAGAGAAAGTAACATCAGAAGTA 90341

Qy         22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
              |||::::::            |||       ||||||||||||||:::|||||||||
Db      90342 GCTGAATTAGTAAAACAA---GGTAAAAAACCGGGTCTTGCGGTTGTGCTCGTAGGAGAT 90398

Qy         42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
                 :::   |||:::|||||||||||||||||:::|||       |||      |||:::
Db      90399 AACCAAGCCTCTCGTACATATGTTAGAAATAAACAAAAGCGAACAGAAGAAGCAGGAATG 90458

Qy         62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
              ||||||       :::   |||||||||       :::|||::::::::::||||||   :::
Db      90459 AAATCTGTTTTAATTGAACTTCCTGAAACAGTAACAGAAGAAAAATTACTAAAAGTAGTA 90518

Qy         82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
                       |||   |||   ::::::|||||||||||||||||||||||||||||||||
Db      90519 GAAGAGTTAAATGAAGATAATACAATTCATGGTATATTGGTACAATTACCTTTACCAAAA 90578

Qy        102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
              |||:::::::::|||   :::::::::|||:::|||   :::|||||||||||||||||
Db      90579 CATATTTCAGAAGAAAAAGTCATTGATGCTATTAGTTTTGACAAAGATGTTGACGGTTTC 90638

Qy        122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
              ||||||:::|||:::|||   |||   :::   |||:::|||       ||||||||||||
Db      90639 CACCCGGTGAATGTAGGTAACTTATTCATT---GGTAAAGATTCT---TTCGTTCCGTGC 90692

Qy        142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
              ||||||   |||   :::|||||:::          |||   :::|||:::||||||||||
Db      90693 ACACCCGCAGGAATTATTGAGTTAATTAAATCAACCGGCACTCAAATAGAAGGCAAACGA 90752

Qy        162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLys 181
              ||||||||||||||||||||||||||||||||   |||   |||   ||||||
Db      90753 GCAGTCGTTATTGGCAGAAGTAATATTGTAGGTAAGCCAGTAGCACAATTATTATTAAAT 90812

Qy        182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
                   ||||||||||||:::|||   ||||||   ||||||          ::::::   :::
```

```
Db      90813 GAAAATGCTACAGTAACAATAGCTCATAGCCGTACAAAAGATTTACCTCAAGTGGCAAAA 90872

Qy        202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
              :::||||||||||::::::  |||  |||:::|||      |||:::     :::|||
Db      90873 GAAGCCGATATTCTAGTTGTTGCAACAGGATTAGCTAAATTTGTTAAAAAAGAATACATT 90932

Qy        222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerPro 241
              ||||||||||  :::|||||||||||:::                 :::
Db      90933 AAGCCAGGCGCAATTGTTATTGATGTTGGTATG----------------GACCGCGAT 90974

Qy        242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
                  :::|||   |||||||||   :::::::::      :::   ||||||
Db      90975 GAAAACAATAAATTATGTGGCGATGTAGATTTTGATGATGTCAAAGAGCAAGCAGGATTC 91034

Qy        262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
              ||||||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db      91035 ATCACTCCAGTTCCAGGAGGCGTTGGGCCAATGACAATTACCATGCTTCTTGCTAATACA 91094

Qy        282 LeuGluSerAlaLysArgIleHisLys 290
              |||::::::|||||||||||||  |||
Db      91095 TTAAAAGCTGCAAAACGCATTTGGAAG 91121

RESULT 11
ABQ67195_3/c
Continuation (4 of 5) of ABQ67195 from base 300001 (Listeria innocua contig DNA sequence
WP Sequence split into 5 fragments  LOCUS ABQ67195 Accession Abq67195
WP      Fragment Name       Begin      End
WP      ABQ67195_0              1      110000
WP      ABQ67195_1         100001      210000
WP      ABQ67195_2         200001      310000
WP      ABQ67195_3         300001      410000
WP      ABQ67195_4         400001      495269

Alignment Scores:
Pred. No.:              7.84e-72        Length:         110000
Score:                  763.50          Matches:        151
Percent Similarity:     72.66%          Conservative:   59
Best Local Similarity:  52.25%          Mismatches:     70
Query Match:            51.38%          Indels:         9
DB:                     6               Gaps:           4

US-10-723-061-18 (1-292) x ABQ67195_3 (1-110000)

Qy          2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
              :::|||||||||||||||   :::||||||:::|||:::|||:::::      |||:::
Db      68669 GAAATTATTGATGGCAAAAAGTTAGCCAAAGAAATCCAAGAGAAAGTAACATCAGAAGTA 68610

Qy         22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
              |||::::::          |||       |||||||||||||||:::||||||||||
Db      68609 GCTGAATTAGTAAAACAA---GGTAAAAAACCGGGTCTTGCGGTTGTGCTCGTAGGAGAT 68553

Qy         42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
                 :::   |||:::|||||||||||||||||||:::|||     |||        |||:::
Db      68552 AACCAAGCCTCTCGTACATATGTTAGAAATAAACAAAAGCGAACAGAAGAAGCAGGAATG 68493

Qy         62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
              ||||||      :::   |||||||||      :::|||:::::::::||||||   :::
Db      68492 AAATCTGTTTTAATTGAACTTCCTGAAACAGTAACAGAAGAAAAATTACTAAAAGTAGTA 68433

Qy         82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
                    |||   |||   ::::::|||||||||||||||||||||||||||||||
Db      68432 GAAGAGTTAAATGAAGATAATACAATTCATGGTATATTGGTACAATTACCTTTACCAAAA 68373

Qy        102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
              |||::::::::|||   :::::::::|||:::|||   :::|||||||||||||||||
Db      68372 CATATTTCAGAAGAAAAAGTCATTGATGCTATTAGTTTTGACAAAGATGTTGACGGTTTC 68313

Qy        122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
              ||||||:::|||:::|||   |||   :::   |||:::|||      ||||||||||||
Db      68312 CACCCGGTGAATGTAGGTAACTTATTCATT---GGTAAAGATTCT---TTCGTTCCGTGC 68259

Qy        142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
              ||||||   |||   :::|||||||:::         |||   :::|||:::||||||||||

Db      68258 ACACCCGCAGGAATTATTGAGTTAATTAAATCAACCGGCACTCAAATAGAAGGCAAACGA 68199

Qy        162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLys 181
              |||||||||||||||||||||||||||||||||||||||||   |||   |||   ||||||
Db      68198 GCAGTCGTTATTGGCAGAAGTAATATTGTAGGTAAGCCAGTAGCACAATTATTATTAAAT 68139

Qy        182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
                    ||||||||||||:::|||   ||||||   ||||||         ::::::   :::
Db      68138 GAAAATGCTACAGTAACAATAGCTCATAGCCGTACAAAAGATTTACCTCAAGTGGCAAAA 68079

Qy        202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
              :::||||||||||::::::  |||  |||:::|||      |||:::     :::|||
Db      68078 GAAGCCGATATTCTAGTTGTTGCAACAGGATTAGCTAAATTTGTTAAAAAAGAATACATT 68019

Qy        222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerPro 241
              ||||||||||  :::|||||||||||:::                 :::
Db      68018 AAGCCAGGCGCAATTGTTATTGATGTTGGTATG----------------GACCGCGAT 67977

Qy        242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
                  :::|||   |||||||||   :::::::::      :::   ||||||
Db      67976 GAAAACAATAAATTATGTGGCGATGTAGATTTTGATGATGTCAAAGAGCAAGCAGGATTC 67917

Qy        262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
              ||||||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db      67916 ATCACTCCAGTTCCAGGAGGCGTTGGGCCAATGACAATTACCATGCTTCTTGCTAATACA 67857

Qy        282 LeuGluSerAlaLysArgIleHisLys 290
              |||::::::|||||||||||||  |||
Db      67856 TTAAAAGCTGCAAAACGCATTTGGAAG 67830

RESULT 12
ABL07155
ID   ABL07155 standard; cDNA; 3003 BP.
XX
AC   ABL07155;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 15947.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB63052.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 15947; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
```

```
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3003 BP; 698 A; 798 C; 889 G; 618 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              7.58e-74        Length:         3003
Score:                  762.00          Matches:        160
Percent Similarity:     68.28%          Conservative:   38
Best Local Similarity:  55.17%          Mismatches:     88
Query Match:            51.28%          Indels:         4
DB:                     4               Gaps:           3

US-10-723-061-18 (1-292) x ABL07155 (1-3003)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||||||   |||      |||||||||  ||||||||||||:::      |||
Db        208 GCCAAGATCATATCGGGCACGGCGGTGGCCAAATCCATTCGCGAAGAGCTGCGTAACGAG 267

Qy         21 IleAlaLysMet---LysAspAlaIleGlyValValProGlyLeuAlaValIleLeuVal 39
              :::      |||   |||                 ||||||||||||   ::::::   |||
Db        268 GTGACGGCTATGAGCAAGCAATTGGCGGATTTTGTGCCCGGCTTAAGGATTGTCCAGGTT 327

Qy         40 GlySerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaVal 59
              |||   |||:::||||||      |||:::|||   |||   ||||||          :::
Db        328 GGTGGACGTGAGGACTCCAACGTTTACATCCGCATGAAGATCAAGGCGGCTACGGAGATC 387

Qy         60 GlyIleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLys 79
              ||||||   :::      |||   ||||||     :::|||   |||:::|||
Db        388 GGTATCGATGCCGCCCACGTCCAGCTGCCCCGATCCATCACCGAGGTGGAATTGCTCGAT 447

Qy         80 HisIleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeu 99
              |||  ·    |||  ||||||  ||||||||||||:::||||||:::||||||
Db        448 AAGATAAACGATCTGAACGAGGACCCCCGGGTGCACGGCATCATAGTGCAAATGCCCCTG 507

Qy        100 ------ProHisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAsp 117
                    ::::::          |||  :::||||||||||  ||||||||||
Db        508 GATTGTGACACACCCATCGATTCGCACCGAATTACGGACGCCGTTTCCCCGGAGAAGGAT 567

Qy        118 ValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhe 137
              |||||||||   |||  :::|||  ||||||||||||:::  |||
Db        568 GTTGACGGCTTGCACACGGTTAACGAAGGTCGCCTGGCCATC---GGTGACCTTGGTGGT 624

Qy        138 PheValProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIle 157
              |||:::||||||||||   ||||||:::||||||:::   |||  ||||||||||||
Db        625 TTCCTGCCCTGCACTCCTTGGGGTTGCCTGGAGTTAATTCGCCGTTCTGGAGTAGAGATC 684

Qy        158 LysGlyLysArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeu 177
                 |||  ||||||||||||:::|||||||||  |||||||||  |||||||||
Db        685 GCCGGAGCCAGGGCTGTGGTTCTGGGACGCAGCAAGATCGTGGGCACTCCCGCCGCCGAA 744

Qy        178 LeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGlu 197
              ||||||:::   ||||||||||||||:::::   ||||||   |||:::      |||
Db        745 CTGCTCAAGTGGGCCAATGCCACGGTCACAGTGTGCCACTCAAAGACCCGCAACCTGGAG 804

Qy        198 GluIleThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArg 217
              ||||||||||   |||||||:::::      :::||||||||||   :::|||:::
Db        805 GAGATCACCCGAAGTGCAGACATTCTGGTGGTCGGAATTGGAGTGGCCGAGATGGTTAAG 864

Qy        218 GlySerTrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAsp 237
              ||||||||||||||||||||||||   :::||||||   |||||||||          |||
Db        865 GGATCGTGGATAAAGCCCGGAGCAGTGGTCATCGATTGCGGCATCAATGTGAAGCCAGAC 924

Qy        238 ProGluSerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLys 257
                       |||   :::||||||||||||||||||   |||   ||||||   :::
Db        925 GCCAGCAAGGCCAGTGGCAGCAAGCTGGTCGGGGACGTGGACTACGCGGAAGCTCTCCAA 984

Qy        258 IleAlaGlyLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeu 277
              :::||||||   :::||||||||||||||||||||||||||||||||:::||||||||||
Db        985 GTGGCAGGACATTTGACGCCAGTGCCCGGCGGCGTGGGACCCATGACGGTGGCCATGCTT 1044

Qy        278 LeuSerAsnThrLeuGluSerAlaLysArg 287
                 :::   ||||||:::   ||||||   |||
Db       1045 ATGAAAAACACCGTGCGATCTGCGGCCCGA 1074

RESULT 13
ABA03041_13
Continuation (14 of 30) of ABA03041 from base 1300001 (Listeria monocytogenes EGD-e genon
WP Sequence split into 30 fragments  LOCUS ABA03041 Accession Aba03041
WP      Fragment Name       Begin      End
WP      ABA03041_00             1    110000
WP      ABA03041_01        100001    210000
WP      ABA03041_02        200001    310000
WP      ABA03041_03        300001    410000
WP      ABA03041_04        400001    510000
WP      ABA03041_05        500001    610000
WP      ABA03041_06        600001    710000
WP      ABA03041_07        700001    810000
WP      ABA03041_08        800001    910000
WP      ABA03041_09        900001   1010000
WP      ABA03041_10       1000001   1110000
WP      ABA03041_11       1100001   1210000
WP      ABA03041_12       1200001   1310000
WP      ABA03041_13       1300001   1410000
WP      ABA03041_14       1400001   1510000
WP      ABA03041_15       1500001   1610000
WP      ABA03041_16       1600001   1710000
WP      ABA03041_17       1700001   1810000
WP      ABA03041_18       1800001   1910000
WP      ABA03041_19       1900001   2010000
WP      ABA03041_20       2000001   2110000
WP      ABA03041_21       2100001   2210000
WP      ABA03041_22       2200001   2310000
WP      ABA03041_23       2300001   2410000
WP      ABA03041_24       2400001   2510000
WP      ABA03041_25       2500001   2610000
WP      ABA03041_26       2600001   2710000
WP      ABA03041_27       2700001   2810000
WP      ABA03041_28       2800001   2910000
WP      ABA03041_29       2900001   2944528

Alignment Scores:
Pred. No.:              6.04e-71        Length:         110000
Score:                  755.50          Matches:        149
Percent Similarity:     72.32%          Conservative:   60
Best Local Similarity:  51.56%          Mismatches:     71
Query Match:            50.84%          Indels:         9
DB:                     6               Gaps:           4

US-10-723-061-18 (1-292) x ABA03041_13 (1-110000)

Qy          2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
              ::||||||||||||||||   :::||||||:::|||:::|||::::::      |||:::
Db      83495 GAAATTATTGATGGCAAAAAGTTAGCAAAAGAAATTCAAGAAAAAGTAACAAGAGAAGTA 83554

Qy         22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
              |||::::::            |||      ||||||||||||||||:::||||||||||
Db      83555 GCTGAATTAGTAAAAGAA---GGTAAACAACCTGGTCTTGCTGTTGTGCTCGTAGGCGAC 83611

Qy         42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
                 :::   |||:::|||||||||||||||||:::|||   |||      |||:::
Db      83612 AATCAAGCATCTCGTACATATGTAAGAAATAAACAAAAAAGAACAGAAGAAGCGGGTATG 83671

Qy         62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
              ||||||      :::   |||||||||:::   :::|||:::::::::|||      :::
Db      83672 AAATCTGTTTTAATTGAACTTCCAGAAAATGTAACAGAAGAAAAATTACTATCTGTCGTA 83731

Qy         82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
                    |||   |||   ::::::||||||||||||||||||||||||||||||||||
Db      83732 GAAGAGCTAAATGAAGATAAAACCATTCACGGCATACTCGTTCAGTTACCACTACCAGAA 83791

Qy        102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
              |||:::::::::|||   :::::::::   :::|||   :::||||||||||||||||||
Db      83792 CACATTTCAGAAGAAAAAGTAATTGATACTATTAGCTTTGACAAAGATGTTGACGGTTTT 83851
```

```
Qy        122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
              ||||||:::|||:::|||   |||   :::   |||:::|||      ||||||||||||
Db      83852 CATCCAGTGAATGTAGGTAATTTATTCATT---GGAAAAGATTCA---TTTGTTCCTTGT 83905

Qy        142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
              ||||||   |||   :::||||||:::         |||   :::|||:::||||||||
Db      83906 ACACCAGCTGGAATTATTGAACTTATAAAATCAACCGGCACTCAAATAGAAGGTAAACGC 83965

Qy        162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLys 181
              ||||||||||||||||||||||||||||||||    |||   |||   ||||||
Db      83966 GCTGTCGTTATTGGTAGAAGTAATATCGTAGGGAAACCAGTAGCCCAATTACTGTTAAAT 84025

Qy        182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
              ||||||||||||:::|||   ||||||   ||||||         ::::::   :::
Db      84026 GAAAACGCGACAGTAACCATTGCGCATAGCCGTACAAAAGATTTACCACAAGTAGCGAAA 84085

Qy        202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
              :::||||||||||::::::   |||   |||:::|||      |||:::      :::|||
Db      84086 GAAGCGGATATTCTTGTTGTAGCAACAGGTTTAGCTAAATTTGTGAAAAAAGACTATATC 84145

Qy        222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerPro 241
              ||||||||||||   :::|||||||||||:::                  :::
Db      84146 AAACCAGGTGCGATTGTTATTGATGTTGGCATG-----------------GATCGCGAC 84187

Qy        242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
                    :::|||   |||||||||   :::::::::      :::   ||||||
Db      84188 GAAAATAATAAGTTATGCGGTGATGTTGACTTTGATGACGTAGTACAAGAAGCGGGATTC 84247

Qy        262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
              ||||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db      84248 ATTACGCCAGTACCAGGTGGCGTTGGCCCGATGACAATCACAATGCTACTTGCGAACACA 84307

Qy        282 LeuGluSerAlaLysArgIleHisLys 290
              |||::::::||||||||||||   |||
Db      84308 TTAAAAGCAGCAAAACGCATTTGGAAA 84334
```

RESULT 14
ABQ71010
ID ABQ71010 standard; DNA; 5635 BP.
XX
AC ABQ71010;
XX
DT 29-AUG-2003 (revised)
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b contig DNA sequence #952.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes ATCC 19115.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR003061.
XX
PR 04-OCT-2000; 2000FR-00012697.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators.
XX
PS Claim 14; SEQ ID NO 3823; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences (ABQ67188-
CC ABQ71212) from Listeria sp. The sequences are useful as probes and
CC primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of gene
CC expression. Proteins encoded by the nucleic acid sequences can be used to
CC screen for compounds that modulate gene expression, replication and
CC pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in anti-
CC Listeria vaccines. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
CC on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 5635 BP; 1896 A; 927 C; 1238 G; 1574 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.66e-72	Length:	5635
Score:	751.50	Matches:	148
Percent Similarity:	73.01%	Conservative:	63
Best Local Similarity:	51.21%	Mismatches:	69
Query Match:	50.57%	Indels:	9
DB:	6	Gaps:	4

US-10-723-061-18 (1-292) x ABQ71010 (1-5635)

```
Qy          2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
              :::|||||||||||||||   :::|||||:::|||:|||::::::      |||:::
Db        439 GAAATTATTGATGGCAAAAAGTTAGCAAAAGAAATTCAAGAAAAAGTAACAAGAGAAGTA 498

Qy         22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
              |||::::::          |||      ||||||||||||||:::|||||||||
Db        499 GCTGAATTAGTAAAAGAA---GGTAAGAAACCAGGGCTTGCTGTTGTGCTCGTTGGCGAC 555

Qy         42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
                 :::   |||:::|||||||||||||||||:::|||      |||      |||:::
Db        556 AATCAAGCATCTCGTACATATGTAAGAAATAAACAAAAACGGACAGAAGAAGCGGGGATG 615

Qy         62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
              ||||||      :::   |||||||||:::   :::|||::::::::|||      :::
Db        616 AAATCCGTTTTAATTGAACTTCCAGAAAATGTAACAGAAGAAAAATTACTATCTGTTGTA 675

Qy         82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
                    |||   |||   ::::::|||||||||||||||||||||||||||||
Db        676 GAGGAGCTTAACGAAGATAAAACTATTCATGGCATACTCGTGCAGTTACCACTACCAGAA 735

Qy        102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
              |||::::::::::|||   :::::::::   :::|||   :::|||||||||||||||||
Db        736 CACATTTCAGAAGAAAAAGTAATTGATACTATTAGCTATGACAAAGATGTTGACGGTTTC 795

Qy        122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
              ||||||:::|||:::|||   |||   :::   |||:::|||      ||||||||||||
Db        796 CATCCAGTGAATGTAGGTAATTTATTCATC---GGAAAAGATTCT---TTTGTTCCTTGT 849

Qy        142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
              ||||||   |||   :::||||||:::         |||   :::|||:::||||||||
Db        850 ACACCAGCAGGAATTATTGAACTTATAAAATCGACCGGCACTCAAATAGAAGGCAAACGC 909

Qy        162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLys 181
              ||||||||||||||||||||||||||||||||    |||   |||   ||||||
Db        910 GCTGTTGTTATTGGTAGAAGTAACATCGTAGGGAAACCAGTAGCTCAATTACTGTTAAAT 969

Qy        182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
              ||||||||||||:::|||   ||||||   ||||||         ::::::   :::
Db        970 GAAAACGCGACAGTAACCATTGCGCATAGCCGTACAAAAGATTTACCTCAAGTAGCGAAA 1029

Qy        202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
              :::||||||||||::::::   |||   |||:::|||      |||:::      :::|||
Db       1030 GAAGCAGATATTCTTGTTGTAGCGACAGGCCTAGCGAAATTCGTGAAAAAAGACTATATC 1089

Qy        222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerPro 241
              ||||||||||||   :::|||:::||||||::::::      ::::::
Db       1090 AAACCAGGTGCGGTTGTTATTAATGTTGGCATGGATCTCGCCGAAAAT------------ 1137
```

```
Qy        242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
                  :::|||   |||||||||   :::::::::      :::   ||||||
Db       1138 ------AATAAGTTATGCGGTGATGTGGACTTTGATGATGTGGTAGAAGAAGCAGGATTT 1191

Qy        262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
              |||||||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db       1192 ATCACGCCGGTACCAGGTGGCGTTGGCCCGATGACAATCACAATGCTACTTGCGAACACA 1251

Qy        282 LeuGluSerAlaLysArgIleHisLys 290
              |||::::::|||||||||||||   |||
Db       1252 TTAAAAGCAGCAAAACGCATTTGGAAA 1278
```

```
RESULT 15
ABN66178
ID   ABN66178 standard; DNA; 858 BP.
XX
AC   ABN66178;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 269.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus pyogenes.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB004789.
XX
PR   27-OCT-2000; 2000GB-00026333.
PR   24-NOV-2000; 2000GB-00028727.
PR   07-MAR-2001; 2001GB-00005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J,  Masignani V,  Margarit Y RosI,  Grandi G,  Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP25547.
XX
PT   New Streptococcus protein for the treatment or prevention of infection or
PT   disease caused by Streptococcus bacteria, such as meningitis, and for
PT   detecting a compound that binds to the protein.
XX
PS   Claim 7; Page 3181-3182; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins
XX
SQ   Sequence 858 BP; 279 A; 153 C; 189 G; 237 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:              1.48e-72        Length:         858
Score:                  743.50          Matches:        149
Percent Similarity:     70.34%          Conservative:   55
Best Local Similarity:  51.38%          Mismatches:     77
Query Match:            50.03%          Indels:         9
DB:                     6               Gaps:           4

US-10-723-061-18 (1-292) x ABN66178 (1-858)

Qy          2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
              ::::::|||||||||||||   :::|||:::::::::::::::|||:::|||   ::::::
Db         13 GAACTAATTGATGGTAAAGCCTTAGCTCAAAAGATGCAACAAGAGTTAGCAGCTAAAGTC 72

Qy         22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
                    :::|||        |||:::|||||||||||||||||||||||||||||||
Db         73 AACAATCTAAAACAAAAAAAAGGAATTGTACCAGGCTTAGCCGTTATTCTTGTAGGTGAT 132

Qy         42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
                      ||||||   ||||||||||||||:::::|||          ||||||
Db        133 GATCCTGCTAGTCAGGTGTATGTCCGTAATAAAGAGCGTGCAGCTCTTACTGTAGGTTTT 192

Qy         62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
              ||||||       |||   |||   |||          :::::::|::::::      |||
Db        193 AAAAGTGAGACCGTTAGATTATCAGAATTCATTTGTCAAGAAGAGCTTATTGCAGTAATC 252

Qy         82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
                    :::|||:::|||   :::::::||||||||||||||||||||||||||||:::
Db        253 GAACGTTACAATGCAGATAACACTATTCATGGTATTTTAGTGCAGCTACCCCTGCCAAAT 312

Qy        102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
              |||:::|||||:::   |||:::   |||:::       :::|||||||||||||||||
Db        313 CATATTAATGATAAAAAAATTATTCTCGCCATTGATCCCAAAAAAGATGTGGATGGTTTT 372

Qy        122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
              ||||||:::|||   |||   |||          ||||||   |||       |||||||||
Db        373 CACCCGATGAATACAGGTCACCTT---TGGTCAGGACGT---CCCTTGATGGTTCCTTGT 426

Qy        142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
              ||||||   |||   ||||||||||||          |||   |||   :::::::||||
Db        427 ACTCCATCAGGGATTATGGAATTGCTTCGAGAATATAATGTTAACCTTGAAGGTAAACAT 486

Qy        162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLys 181
              ||||||:::|||||||||||||||||||||||   |||   |||   ||||||
Db        487 GCCGTCATTATTGGCAGATCGAATATCGTTGGAAAACCAATGGCACAGCTCTTACTGGAC 546

Qy        182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
              ||||||||||||:::::   ||||||   |||::::::   ||||||:::   |||
Db        547 AAAAATGCAACAGTCACGTTGACACATTCAAGAACACGTCAATTAGAAGAAGTATGTCGC 606

Qy        202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
              ||||||::::::|||   |||:::|||      :::   :::         :::|||
Db        607 TGTGCAGATGTGTTGATTGTGGCAATCGGACAAGGTCATTTCATAACAAAACAATATATA 666

Qy        222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerPro 241
              |||   ||||||   :::|||||||||||:::|||       |||:::
Db        667 AAAGATGGTGCGATAGTGATTGATGTAGGAATGAACCGTGATGATAATGGC--------- 717

Qy        242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
                        :::|||:::|||||||||   :::::::|||   :::::::|||
Db        718 ---------AAGCTAATTGGAGATGTGGCCTTTGATGAGGTGGCAGAAGTTGCAGCGAAA 768

Qy        262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        769 ATCACCCCTGTACCAGGAGGTGTCGGTCCTATGACGATTGCTATGTTGCTAGAGCAAACT 828

Qy        282 LeuGluSerAlaLysArg---IleHisLys 290
                 :::||||||   |||      ||||||
Db        829 TATCAATCTGCTCTCCGTAGTACTCATAAA 858
```

Search completed: July 24, 2004, 07:19:50

Job time : 625 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 24, 2004, 06:52:51 ; Search time 91 Seconds
(without alignments)
1780.722 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10723061/runat_21072004_090401_7143/app_query.fasta_1.455
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10723061_@CGN_1_1_128_@runat_21072004_090401_7143 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1030	69.3	926	4	US-09-903-814A-19	Sequence 19, Appl
	2	1016	68.4	1076	4	US-09-903-814A-15	Sequence 15, Appl
	3	740.5	49.8	875	4	US-09-903-814A-21	Sequence 21, Appl
	4	735	49.5	3112	3	US-09-318-448-9	Sequence 9, Appli
	5	724	48.7	13440	4	US-08-961-527-128	Sequence 128, App
	6	712.5	47.9	855	4	US-09-107-532A-3634	Sequence 3634, Ap
	7	684	46.0	957	4	US-09-489-039A-6086	Sequence 6086, Ap
c	8	681.5	45.9	1830121	4	US-09-557-884-1	Sequence 1, Appli
c	9	681.5	45.9	1830121	4	US-09-643-990A-1	Sequence 1, Appli
	10	679.5	45.7	873	4	US-09-134-001C-564	Sequence 564, App
	11	677	45.6	957	4	US-09-252-991A-13712	Sequence 13712, A
	12	677	45.6	1188	4	US-09-252-991A-13926	Sequence 13926, A
	13	670.5	45.1	9062	4	US-08-956-171E-85	Sequence 85, Appl
	14	663.5	44.7	840	4	US-09-134-000C-460	Sequence 460, App
	15	661	44.5	873	4	US-09-543-681A-2906	Sequence 2906, Ap
	16	638	42.9	1371	4	US-08-858-207A-89	Sequence 89, Appl
	17	608.5	40.9	945	4	US-09-328-352-1366	Sequence 1366, Ap
	18	590	39.7	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	19	589	39.6	640681	4	US-09-790-988-1	Sequence 1, Appli
c	20	584.5	39.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	21	584.5	39.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	22	544.5	36.6	579	4	US-09-903-814A-17	Sequence 17, Appl
	23	541	36.4	956	4	US-09-221-017B-484	Sequence 484, App
c	24	464	31.2	600	4	US-09-252-991A-13432	Sequence 13432, A
c	25	385	25.9	669	4	US-09-252-991A-13532	Sequence 13532, A
	26	356.5	24.0	578	3	US-09-328-111-765	Sequence 765, App
c	27	355	23.9	546	4	US-09-252-991A-13632	Sequence 13632, A
c	28	296	19.9	580073	4	US-08-545-528D-1	Sequence 1, Appli
	29	229.5	15.4	4258	4	US-09-596-002-5	Sequence 5, Appli
c	30	213	14.3	408	4	US-09-252-991A-13433	Sequence 13433, A
c	31	211	14.2	338	4	US-08-651-155B-33	Sequence 33, Appl
c	32	211	14.2	338	4	US-09-194-036B-33	Sequence 33, Appl
c	33	198	13.3	429	4	US-09-596-002-1	Sequence 1, Appli
c	34	104.5	7.0	1428	4	US-09-252-991A-2357	Sequence 2357, Ap
	35	104.5	7.0	1449	4	US-09-252-991A-2559	Sequence 2559, Ap
c	36	104.5	7.0	2343	4	US-09-252-991A-2275	Sequence 2275, Ap
c	37	103	6.9	9280	4	US-08-956-171E-131	Sequence 131, App
c	38	100	6.7	246	4	US-09-543-681A-2795	Sequence 2795, Ap
	39	99.5	6.7	696	4	US-08-426-630-53	Sequence 53, Appl
	40	99.5	6.7	955	4	US-08-426-630-52	Sequence 52, Appl
	41	98	6.6	1149	4	US-09-134-001C-928	Sequence 928, App
c	42	96	6.5	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	43	92	6.2	3078	4	US-09-711-164-272	Sequence 272, App
	44	90.5	6.1	1230	4	US-09-134-001C-953	Sequence 953, App
	45	89	6.0	18475	4	US-08-961-527-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-09-903-814A-19
; Sequence 19, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB1179 USDIV
; CURRENT APPLICATION NUMBER: US/09/903,814A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US 60/092,869
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: US 09/351,703
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 926
; TYPE: DNA
; ORGANISM: Glycine max
US-09-903-814A-19

Alignment Scores:

Pred. No.:	2.42e-119	Length:	926
Score:	1030.00	Matches:	194
Percent Similarity:	89.63%	Conservative:	22
Best Local Similarity:	80.50%	Mismatches:	25
Query Match:	69.31%	Indels:	0
DB:	4	Gaps:	0

US-10-723-061-18 (1-292) x US-09-903-814A-19 (1-926)

```
Qy         52 LysLysLysAlaCysGluAlaValGlyIleLysSerTyrGluValAsnLeuProGluAsp 71
Db          6 GAGAAGAAGGCTTGTGAATCTGTTGGAATCAATTCTTTGGAAGCAAATCTGCCTGAGAAT 65

Qy         72 SerSerGluAspGluValLeuLysHisIleAlaThrPheAsnSerAspProSerValHis 91
Db         66 TCCACAGAAGAAGAAGTTTTGAACTATATTGCAGGCTACAATGATGATCCTTCAGTTCAT 125

Qy         92 GlyIleLeuValGlnLeuProLeuProHisHisMetAsnAspGluAsnIleLeuAsnAla 111
Db        126 GGCATCCTCGTTCAGTTACCCTTACCTTCGCATATGAATGAGCAGAACATCTTGAACGCT 185

Qy        112 ValSerIleGluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMet 131
Db        186 GTCAGGATTGAGAAGGATGTAGATGGTTTTCATCCGTTAAATATTGGTCGTCTTGCCATG 245

Qy        132 GlnGlyArgAspProPhePheValProCysThrProLysGlyCysMetGluLeuLeuHis 151
Db        246 CGTGGAAGAGAACCTCTGTTTGTTCCTTGTACACCAAAGGGATGCATAGAGCTACTGCAC 305

Qy        152 ArgTyrGlyValGluIleLysGlyLysArgAlaValValIleGlyArgSerAsnIleVal 171
Db        306 AGATACAATGTTTCTATTAAAGGAAAGAGGGCTGTTGTGATTGGTCGGAGCAATATTGTT 365

Qy        172 GlyMetProAlaAlaLeuLeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSer 191
Db        366 GGAATGCCAGCTGCTCTCTTGCTTCAAAGGGAAGATGCTACTGTCAGTATTGTCCATTCT 425

Qy        192 AsnThrLysLysProGluGluIleThrArgGlnAlaAspIleValIleAlaAlaValGly 211
Db        426 AGAACCAGTAACCCCGAAGAGATCATAAGACAGGCAGATATTATCATTGCTGCTGTTGGG 485

Qy        212 ValAlaAsnLeuValArgGlySerTrpIleLysProGlyAlaAlaIleIleAspValGly 231
Db        486 CAAGCAAACATGGTGAGGGGAAGCTGGATAAAGCCTGGTGCAGTCATTATTGATGTTGGA 545

Qy        232 IleAsnProValAspAspProGluSerProArgGlyTyrArgLeuValGlyAspValCys 251
Db        546 ATCAACCCGGTAGAGGATCCAAATAGCCCCCGAGGTTACAAACTGGTGGGAGATGTTTGT 605

Qy        252 TyrGluGluAlaSerLysIleAlaGlyLeuIleThrProValProGlyGlyValGlyPro 271
Db        606 TATGAAGAAGCCATAAGAATTGCCTCTGCTGTTACACCAGTTCCTGGAGGAGTTGGTCCA 665

Qy        272 MetThrIleAlaMetLeuLeuSerAsnThrLeuGluSerAlaLysArgIleHisLysPhe 291
Db        666 ATGACCATAGCAATGCTTCTACAAAATACACTCATCTCTGCAAAGAGGGTGCACAATTTT 725

Qy        292 Lys 292
Db        726 GAA 728
```

```
RESULT 2
US-09-903-814A-15
; Sequence 15, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV
;  CURRENT APPLICATION NUMBER: US/09/903,814A
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
;   LENGTH: 1076
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-903-814A-15

Alignment Scores:
Pred. No.:                1.76e-117     Length:          1076
Score:                    1016.00       Matches:         190
Percent Similarity:       81.18%        Conservative:    43
Best Local Similarity:    66.20%        Mismatches:      54
Query Match:              68.37%        Indels:          0
DB:                       4             Gaps:            0

US-10-723-061-18 (1-292) x US-09-903-814A-15 (1-1076)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db         32 GCGCAGATCATCGATGGCAAGGCCATCGCCGCCGACGTCCGCCGCGAGGTCGCCGCCGAT 91

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db         92 GTGGCCGCGCTCTCGTCGGCCCACGGACTCGTGCCGGGGCTGGCCGTGGTCATCGTGGGG 151

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db        152 AGCAGGAAGGACTCGCAGACGTACGTGAACATGAAGCGCAAGGCGTGCGCCGAGGTCGGC 211

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db        212 ATCTGCTCCATCGACGTCGACCTCCCGGAGGACATCTCCGAGACCGCGCTCGTCGCCGAG 271

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db        272 GTTCATCGCCTCAACGCTGACCCCGCAGTGCACGGGATCCTTGTCCAGCTTCCACTTCCT 331

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db        332 AAGCATATCAACGAAGAGAAGATACTGAGCGAGATTTCCATCGAGAAAGATGTGGATGGC 391

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
Db        392 TTCCATCCTCTCAACATTGGCAAGCTTGCAATGAAAGGCAGAGAGCCACTGTTCGTACCA 451

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db        452 TGTACGCCAAAGGGGTGCATGGAGCTCTTGTCAAGGAGCGGAGTCACTGTTAAAGGTAAG 511

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
Db        512 CGGGCAGTTGTGGTTGGTCGCAGCAACATCGTCGGGCTACCTGTATCCCTGCTCCTTCTG 571

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db        572 AAGGCAGATGCGACCGTATCTGTTGTGCACTCGCGGACCCCTGATCCTGAAAGCATTGTA 631

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db        632 CGCGAAGCTGACATAGTCATCGCGGCAGCTGGGCAGGCTATGATGATCAAAGGTGACTGG 691

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db        692 ATCAAGCCAGGTGCTGCGGTCATCGATGTCGGGACGAACTCCATCGATGACCCTACCCGG 751

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db        752 AAGTCCGGGGTACGGCTCGTCGGCGATGTGGATTTCGCAGCGGCGAGCAAGGTTGCTGGG 811

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db        812 TACCTGACTCCGGTTCCCGGAGGCGTTGGCCCAATGACGGTGGCAATGCTGCTGAAGAAC 871

Qy        281 ThrLeuGluSerAlaLysArg 287
```

```
Db        872 ACGGTGGATGGGGCAAAGCGG 892

RESULT 3
US-09-903-814A-21
; Sequence 21, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV
;  CURRENT APPLICATION NUMBER: US/09/903,814A
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
;   LENGTH: 875
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (584)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (680)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (699)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (704)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (756)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (767)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (771)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (780)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (800)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (802)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (816)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (834)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (837)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (842)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (846)..(847)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (858)..(859)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (870)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (874)
;   OTHER INFORMATION: n is a, c, g or t
US-09-903-814A-21

Alignment Scores:
Pred. No.:              4.35e-83        Length:         875
Score:                  740.50          Matches:        146
Percent Similarity:     79.31%          Conservative:   38
Best Local Similarity:  62.93%          Mismatches:     47
Query Match:            49.83%          Indels:         2
DB:                     4               Gaps:           0

US-10-723-061-18 (1-292) x US-09-903-814A-21 (1-875)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||:::|||||||||||||||||   :::|||   :::||||||   ||||||      |||
Db         20 GCGCAAATCATCGACGGCAAGGCCATCGCCGCCGAAATCAGGCGCGAGATCGGCGCCGAG 79

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::|||   :::      |||      :::||||||||||||||||||:::::||||||
Db         80 GTCGCCGTGCTCTCGTCCGCCCACAACATCGTGCCGGGGCTGGCGGTGGTGATCGTGGGG 139

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||:::   |||:::|||||||||      ||||||
Db        140 AGCAGGAAGGACTCGCAGACGTACGTGCAGATGAAGCGCAAGGCCTGCGCCGAGGTCGGC 199

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||:::|||::::::|||:::|||||||||||||   :::|||      ::::::
Db        200 ATCCGCTCCTTCGACGTCGACCTCCCCGAGGACATCGCCGAGGCCGCGCTCGTCGCCGAG 259

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::         |||:::||||||:::||||||||||||||||||||||||||||||||||
Db        260 GTCCACCGCCTCAACGCCGACCCCGCCGTCCACGGAATTCTTGTTCAGCTTCCATTGCCC 319

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||:::|||:::|||||||||||||||   :::||||||||||||||||||||||||
Db        320 AAGCATATCAACGAAGAAAATATCTTAAACCAGATCTCCATTGAGAAAGATGTCGACGGC 379

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||:::|||||||||:::||||||||||||   |||||||||
Db        380 TTTCATCCTTTGAACATTGGCAAGCTTGCAATGAAAGGCAGAGATCCACTGTTCGTACCT 439

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||||||||||||||||   |||   ||||||   :::||||||||||
Db        440 TGCACGCCAAAGGGATGCATGGAGCTCCTGTCACGAAGTGGCGTCACTGTAAAAGGAAAA 499

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAla-AlaLeuLeuLeuGl 180
              |||||||||:::||||||||||||||||||||:::|||::: :::||||||||||
Db        500 CACGCAGTTGTGGTTGGGCGTAGCAACATCGTGGGTTTACCAAGTATCCCTTCTCCTTCT 559
```

```
Qy        180 nLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleTh 200
Db        560 GAAAGCGGACGCTACCGTGTCGATNGTGCATCAACGGACCCAAAT-CCCCAAACAATTTC 618

Qy        200 rArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTr 220
Db        619 CGTCAAGCAAGACATTGTCATTGCAGCAGCTGGGCAAGCCATGATGATCAAGGGAGACTG 678

Qy        220 pIleLysProGlyAlaAlaIleIleAspValGly 231
Db        679 GNTTAAACAAAAGCGCAACGNCATCNACGTCGGG 712

RESULT 4
US-09-318-448-9
; Sequence 9, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
;  APPLICANT: Johnson, William G.
;  APPLICANT: Stenroos, Edward S.
;  TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
;  FILE REFERENCE: 601-1-057
;  CURRENT APPLICATION NUMBER: US/09/318,448
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3112
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-448-9

Alignment Scores:
Pred. No.:              1.67e-81        Length:         3112
Score:                  735.00          Matches:        145
Percent Similarity:     67.91%          Conservative:   56
Best Local Similarity:  48.99%          Mismatches:     91
Query Match:            49.46%          Indels:         4
DB:                     3               Gaps:           3

US-10-723-061-18 (1-292) x US-09-318-448-9 (1-3112)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db         63 GCAGAAATCCTGAACGGGAAGGAGATCTCCGCGCAAATAAGGGCGAGACTGAAAAATCAA 122

Qy         21 IleAlaLysMetLysAspAlaIle---GlyValValProGlyLeuAlaValIleLeuVal 39
Db        123 GTCACTCAGTTGAAGGAGCAAGTACCTGGTTTCACACCACGCCTGGCAATATTACAGGTT 182

Qy         40 GlySerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaVal 59
Db        183 GGCAACAGAGATGATTCCAATCTTTATATAAATGTGAAGCTGAAGGCTGCTGAAGAGATT 242

Qy         60 GlyIleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLys 79
Db        243 GGGATCAAAGCCACTCACATTAAGTTACCAAGAACAACCACAGAATCTGAGGTGATGAAG 302

Qy         80 HisIleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeu 99
Db        303 TACATTACATCTTTGAATGAAGACTCTACTGTACATGGGTTCTTAGTGCAGCTACCTTTA 362

Qy        100 ------ProHisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAsp 117
Db        363 GATTCAGAGAATTCCATTAACACTGAAGAAGTGATCAATGCTATTGCACCCGAGAAGGAT 422

Qy        118 ValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhe 137
Db        423 GTGGATGGATTGACTAGCATCAATGCTGGGAGACTTGCTAGAGGTGACCTCAATGACTGT 482

Qy        138 PheValProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIle 157
Db        483 TTCATTCCTTGTACGCCTAAGGGATGCTTGGAACTCATCAAAGAGACAGGGGTGCCGATT 542

Qy        158 LysGlyLysArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeu 177
Db        543 GCCGGAAGGCATGCTGTGGTGGTTGGGCGCAGTAAAATAGTTGGGGCCCCGATGCATGAC 602

Qy        178 LeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGlu 197
Db        603 TTGCTTCTGTGGAACAATGCCACAGTGACCACCTGCCACTCCAAGACTGCCCATCTGGAT 662

Qy        198 GluIleThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArg 217
Db        663 GAGGAGGTAAATAAAGGTGACATCCTGGTGGTTGCAACTGGTCAGCCTGAAATGGTTAAA 722

Qy        218 GlySerTrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAsp 237
Db        723 GGGGAGTGGATCAAACCTGGGGCAATAGTCATCGACTGTGGAATCAATTATGTCCCAGAT 782

Qy        238 ProGluSerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLys 257
Db        783 GATAAAAAACCAAATGGGAGAAAAGTTGTGGGTGATGTGGCATACGACGAGGCCAAAGAG 842

Qy        258 IleAlaGlyLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeu 277
Db        843 AGGGCGAGCTTCATCACTCCTGTTCCTGGCGGCGTAGGGCCCATGACAGTTGCAATGCTC 902

Qy        278 LeuSerAsnThrLeuGluSerAlaLysArg---IleHisLysPheLys 292
Db        903 ATGCAGAGCACAGTAGAGAGTGCCAAGCGTTTCCTGGAGAAATTTAAG 950

RESULT 5
US-08-961-527-128
; Sequence 128, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  128:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13440 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-128

Alignment Scores:
Pred. No.:              4.3e-79         Length:         13440
```

```
Score:                  724.00          Matches:        143
Percent Similarity:     68.88%          Conservative:   54
Best Local Similarity:  50.00%          Mismatches:     81
Query Match:            48.72%          Indels:         8
DB:                     4               Gaps:           2

US-10-723-061-18 (1-292) x US-08-961-527-128 (1-13440)

Qy          2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
              :::|||||||||||||||||   :::|||   :::::::::   :::::::|||   :::
Db       1558 CAGATTATTGATGGGAAAGCTTTAGCGGCCAAATTGCAGGGGCAGTTGGCTGAAAAGACT 1617

Qy         22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
              ||||||:::|||:::      |||:::|||||||||||||   ||||||||||||||
Db       1618 GCAAAATTAAAGGAAGAAACAGGTCTAGTGCCTGGTTTGGTAGTGATTTTGGTTGGGGAC 1677

Qy         42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
                  ||||||   ||||||||||||||||:::::::::      |||   |||
Db       1678 AATCCAGCCAGCCAAGTCTACGTTCGCAACAAGGAGAGGTCAGCCCTTGCGGCTGGTTTC 1737

Qy         62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
              :::|||      |||   :::||||||      :::::::::|||:::|||      |||
Db       1738 CGTAGCGAAGTAGTACGGGTTCCAGAGACCATTACTCAAGAGGAATTGTTAGACCTGATT 1797

Qy         82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
              |||   :::|||   ||||||:::   ||||||||||||||||||||||||||||||||
Db       1798 GCTAAATACAATCAGGATCCAGCTTGGCATGGGATTTTGGTTCAGTTGCCATTACCAAAA 1857

Qy        102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
              |||::::::::::|||   :::|||   |||:::      ||||||||||||||||||
Db       1858 CACATTGATGAAGAGGCGGTTCTATTGGCTATTGACCCAGAAAAGGATGTGGATGGTTTC 1917

Qy        122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
              ||||||||||||:::|||||||||                |||       :::|||
Db       1918 CATCCTCTAAACATGGGGCGTCTT------TGGTCTGGTCATCCAGTCATGATTCCTTCG 1971

Qy        142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
              ||||||   |||   ||||||:::   |||   ||||||:::::::::::|||||||
Db       1972 ACACCGGCAGGAATTATGGAAATGTTCCATGAATATGGGATTGACTTGGAAGGTAAAAAT 2031

Qy        162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLys 181
              |||||||||||||||||||||||||||||||||   |||   |||   ||||||
Db       2032 GCAGTCGTCATCGGTCGATCCAATATTGTCGGAAAACCTATGGCCCAGCTTCTTTTGGCA 2091

Qy        182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
                 ||||||||||:::::   ||||||   |||         ::::::
Db       2092 AAGAATGCAACAGTAACCTTGACTCACTCACGTACTCATAATCTTTCCAAGGTGGCTGCA 2151

Qy        202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
              :::|||||||||:::::   |||:::|||   |||       |||         ::::::
Db       2152 AAAGCAGATATTCTGGTTGTTGCAATCGGTCGTGCCAAGTTTGTGACTGCTGACTTTGTC 2211

Qy        222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerPro 241
              ||||||||||||   :::|||||||||||||:::|||      ::::::
Db       2212 AAACCAGGTGCGGTAGTCATTGACGTTGGGATGAACCGCGATGAAAATGGT--------- 2262

Qy        242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
                 :::|||   ||||||||||   ||||||      :::   :::|||
Db       2263 ---------AAGCTCTGTGGGGATGTTGATTATGAGGCGGTTGCCCCACTTGCTAGCCAC 2313

Qy        262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
              ||||||||||||||||||||||||||||||||||||||   ||||||:::      |||
Db       2314 ATTACGCCAGTCCCTGGAGGTGTCGGTCCTATGACCATTACTATGCTGATGGAGCAAACC 2373

Qy        282 LeuGluSerAlaLysArg 287
                 :::::::|||   |||
Db       2374 TATCAGGCAGCACTTAGG 2391

RESULT 6
US-09-107-532A-3634
; Sequence 3634, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                           ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 7310
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: PC
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/085,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 3634:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 855 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: circular
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Enterococcus faecium
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...855
;       SEQUENCE DESCRIPTION: SEQ ID NO: 3634:
US-09-107-532A-3634

Alignment Scores:
Pred. No.:              1.36e-79        Length:         855
Score:                  712.50          Matches:        146
Percent Similarity:     69.93%          Conservative:   54
Best Local Similarity:  51.05%          Mismatches:     77
Query Match:            47.95%          Indels:         9
DB:                     4               Gaps:           5

US-10-723-061-18 (1-292) x US-09-107-532A-3634 (1-855)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||::::::|||:::||||||   :::|||:::::::::::   |||||||||   :::
Db          4 GCAGAATTGATTAACGGAAAAGAATTAGCTGAAAAAATGCAGGCAGAGATTGCTGAAAAA 63

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              |||   |||:::||||||      |||:::   |||||||||   |||:::|||||||||
Db         64 ATCCAAAAATTGAAAGATAAT---GGTATCCATCCCGGATTAGTTGTTTTATTGGTAGGG 120

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
                     ||||||   |||||||||||||||:::::      :::   :::|||
Db        121 GAAAATCCAGCCAGCCAAATCTATGTAAGAAATAAAGAACGTGTAGCAAAAGAAATCGGA 180

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||   |||            ||||||      ||||||||||:::|||
Db        181 ATCCATTCATTGGTAGAGCGTTATCCTGAGACGATTTCTGAAGACGAACTACTTGCAGAA 240
```

```
Qy          81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
               |||       :::|||  |||||||||  ||||||||||||||||||||||||||||||||
Db         241 ATTGAAAAATACAATCAAGATCCTAGTTTTCATGGTATATTGGTCCAATTACCATTGCCA 300

Qy         101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                  |||:::::::::|||   ||||||   |||:::      ||||||||||||||||||
Db         301 AAGCATATAGATGAAGAAAAGATCCTGTTAGCTATCCGCCCTGAAAAAGACGTAGACGGT 360

Qy         121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
               |||||||||:::|||:::|||||||||   :::   |||:::   |||      :::|||
Db         361 TTTCATCCGATGAATTTAGGCAGACTC---CTTGCGGGTAAA---CCGGACAAGATTCCT 414

Qy         141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
               |||||||||   |||   |||::::::          |||   ::::::      ||||||
Db         415 TGTACTCCATACGGCATCATGAAAATGTTTGAAGCTTATTCTATCGATCCGGCGGGTAAA 474

Qy         161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
               ||||||||||||||||||||||||||||||||||||||   |||   |||   ||||||
Db         475 CGAGCAGTAGTTATCGGACGAAGCAATATCGTCGGTAAACCGATGGCTCAGCTACTCTTG 534

Qy         181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
                  |||:::||||||||||::::::   ||||||   |||:::          |||:::
Db         535 ATGGCAGATGCAACTGTTACTGTTGCTCATTCTAAAACCGAAAATCTGGCTGAACTGGCA 594

Qy         201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
               ::::::||||||||||::::::   |||:::|||      :::   |||          :::
Db         595 CAAGAAGCAGATATTTTAGTCGTAGCAATAGGACGTGGACATTTTGTCACAAAAGAATTC 654

Qy         221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
               :::|||||||||||   :::||||||||||||:::|||
Db         655 GTAAAACCAGGTGCTGTAGTGATAGATGTAGGGATGAAT--------------CGTGAT 699

Qy         241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                  |||   :::|||:::|||||||||   :::::::|||         |||||||||
Db         700 GCTAACGGT---AAGCTGATCGGTGATGTGAAATTTGATGAAGTGGAGCCTATTGCAGGC 756

Qy         261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
               |||||||||||||||   ||||||||||||||||||||||   ||||||:::
Db         757 TGGATCACTCCTGTGCCAAAGGGTGTAGGGCCAATGACTATCACGATGCTGATGTATCAA 816

Qy         281 ThrLeuGluSerAlaLys 286
               |||::::::||||||:::
Db         817 ACCGTAAAATCTGCCGAA 834

RESULT 7
US-09-489-039A-6086
; Sequence 6086, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 6086
;   LENGTH: 957
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6086

Alignment Scores:
Pred. No.:              6.09e-76      Length:         957
Score:                  684.00        Matches:        142
Percent Similarity:     65.52%        Conservative:   48
Best Local Similarity:  48.97%        Mismatches:     90
Query Match:            46.03%        Indels:         10
DB:                     4             Gaps:           4
```

US-10-723-061-18 (1-292) x US-09-489-039A-6086 (1-957)

```
Qy           1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
               ||||||||||||||||||||   :::|||:::|||:::|||   ||:::|||   |||
Db          97 GCAAAAATTATTGACGGTAAAACGATTGCGCAGCAGGTACGCTCTGAGGTTGCG---GAA 153

Qy          21 IleAlaLysMetLysAspAlaIleGlyVal---ValProGlyLeuAlaValIleLeuVal 39
                  |||   :::   |||   |||          |||||||||||||||:::|||||||
Db         154 AAAGTGAAGGCTCGCGTTGCGGCCGGAAAACGCGCCCCTGGGCTGGCCGTCGTGCTGGTC 213

Qy          40 GlySerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaVal 59
               ||||||      ||||||   ||||||   :::|||:::||||||||||||   |||
Db         214 GGCAGCAACCCGGCCTCGCAGATTTATGTCGGCAGCAAGCGCAAAGCATGTGAAGAAGTG 273

Qy          60 GlyIleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLys 79
               |||      |||          :::||||||||   :::|||||   |||:::|||:::
Db         274 GGCTTCGTCTCCCGCTCTTACGATCTCCCGGAAACCACCAGCGAAGCCGAGCTGCTGGAG 333

Qy          80 HisIleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeu 99
                  |||   |||   |||:::|||   ::::::   ||||||||||||||||||||||
Db         334 CTTATCGACACTCTGAATGCCGATAAGACCATCGACGGTATTCTGGTTCAGCTGCCCCTG 393

Qy         100 ProHisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAsp 119
               |||       :::::::::       :::|||       ::::::   :::||||||||||||
Db         394 CCGGCAGGGATCGATAACGTCAAAGTTCTCGAGCGCATCGCGCCGGATAAAGACGTCGAC 453

Qy         120 GlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheVal 139
               ||||||||||||   |||:::|||||||||                |||   |||
Db         454 GGCTTCCATCCTTACAACGTTGGCCGCCTGTGCCAG------CGCGCGCCGCGCCTGCGT 507

Qy         140 ProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGly 159
               ||||||||||||:::|||  :::   ||||||   ||||||   ::::::      |||
Db         508 CCGTGCACTCCACGCGGTATCGTGACCTTGCTGGAACGCTACAATATCGACACCTACGGC 567

Qy         160 LysArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeu 179
                  ||||||||||||||||   ||||||||||||||   |||   ::::::   |||
Db         568 CTCAATGCGGTGGTCATTGGCGCCTCCAATATCGTCGGTCGCCCGATGAGCATGGAGCTG 627

Qy         180 GlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIle 199
                  |||       |||   ::::::   |||       ||||||
Db         628 CTGCTGGCCGGCTGCACCACCACCGTCACCCACCGCTTTACAAAAAACCTGCGCCATCAT 687

Qy         200 ThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySer 219
                  ||||||::::::|||   |||||||||             :::   |||
Db         688 GTCGAAAACGCCGACCTGCTGATCGTCGCGGTGGGCAAACCGGGCTTTATTCCTGGCGAG 747

Qy         220 TrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGlu 239
               |||||||||   ||||||   ::::::||||||||||||||||   ::::::
Db         748 TGGATTAAAGAAGGGGCGATTGTGGTCGATGTCGGCATCAACCGTCTGGAAAGCGGC--- 804

Qy         240 SerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAla 259
                        ::::::||||||||||||   ||||||:::|||::::::   |||
Db         805 --------------AAAGTGGTCGGCGACGTGGTGTATGAAGATGCCGCCGAACGCGCG 849

Qy         260 GlyLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSer 279
                  ||||||||||||||||||||||||||||||||||:::|||   |||:::
Db         850 TCCTACATCACCCCGGTTCCCGGCGGCGTTGGCCCGATGACCGTCGCTACCCTGATTCAG 909

Qy         280 AsnThrLeuGluSerAlaLysArgIleHis 289
               |||||||||::::::   :::      |||
Db         910 AACACGCTGCAGGCGTGCGAAGAGTATCAC 939

RESULT 8
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
;  GENERAL INFORMATION:
;       APPLICANT: Fleischmann et al.
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
```

```
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;          COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;          CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/557,884
;             FILING DATE: 25-Apr-2000
;             CLASSIFICATION: <Unknown>
;          PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/476,102
;             FILING DATE: JUN-5-1995
;          ATTORNEY/AGENT INFORMATION:
;             NAME: Michelle S. Marks
;             REGISTRATION NUMBER: 41,971
;             REFERENCE/DOCKET NUMBER: PB186P3
;          TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-309-8504
;             TELEFAX: 301-309-8439
;       INFORMATION FOR SEQ ID NO: 1:
;          SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;          SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.:                2.68e-70        Length:          1830121
Score:                    681.50          Matches:         144
Percent Similarity:       66.67%          Conservative:    44
Best Local Similarity:    51.06%          Mismatches:      85
Query Match:              45.86%          Indels:          9
DB:                       4               Gaps:            2

US-10-723-061-18 (1-292) x US-09-557-884-1 (1-1830121)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||||   |||    ::::::|||||||||:::     :::|||  :::
Db     640319 GCAAAAATCATCTCTGGTACTGAACTTTCAAAACAAATTAAAGCTAATTTAGCAGATAAA 640260

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              |||          :::          |||||||||||||||||||||||||
Db     640259 ATAACCCATTATATTGAACAAGGAAAACGCGCTCCAGGACTTGCTGTAATCCTCGTTGGT 640200

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              :::          ||||||   ||||||   ||||||:::|||:::||||||   ||||||
Db     640199 GCTGATCCTGCTTCGCAAATTTATGTGGGCAATAAACGTAAAAGTTGCGAGGAAGTCGGT 640140

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||   |||          :::||||||||   ::::::::::::|||:::|||
Db     640139 ATTCTCTCTAAATCTTATGACCTCCCTGAAACAACCACGCAAAATGAATTATTAGCGATT 640080

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              |||          |||:::|||   ::::::   ||||||||||||||||||||||||
Db     640079 ATTGACCAACTTAACGCAGATAAAAATATTGACGGTATTTTGGTGCAACTCCCCTTGCCA 640020

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                 :::|||   |||   |||:::      :::      :::||||||||||||||
Db     640019 AAACAAATTAATGCGGAAGCTATTATTGAACGTATCGATCCTAAAAAAGATGTAGATGGC 639960

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||   |||:::||||||||||             |||   |||
Db     639959 TTTCATCCTTATAATGTTGGGCGGTTATGCCAA------CGTATTCCAACTTTACGTGCT 639906

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||||||||   |||   |||::::|||||||||:::     :::|||  :::
Db     639905 TGTACACCTTATGGTGTGATGAAACTATTAGAAACAACGGGTATTGATTTACACGGCAAG 639846

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||:::::|||   ||||||||||||||   |||   :::|||   |||
Db     639845 CACGCAGTAATTGTAGGCGCATCCAATATTGTCGGCCGCCCAATGTCTTTAGAATTATTA 639786

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              |||   |||||||||:::::   |||       ||||||       |||
Db     639785 TTAGCAGGGGCAACCGTCACAGTAACACATAGATTCACTAAAAATTTGGAAAACCACGTT 639726

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              ||||||||||||||:::::   ||||||||       ||||||:::   |||   |||
Db     639725 CGCCAAGCAGATATACTTGTCGTTGCGGTAGGTAAACCAAATTTAATTTCTGGCGATTGG 639666

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||      |||   :::||||||||||||||||||   ||||||
Db     639665 ATCAAAGAAAGTGCGGTAGTAATTGATGTTGGAATTAACCGAGTAGATGGA--------- 639615

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                       :::|||||||||||||||   :::::::::|||::::::   |||
Db     639614 -----------AAATTAGTAGGCGATGTAGAGTTTGATAAAGCAGCCGAAAAAGCGGCT 639567

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||:::|||||||||:::||||||
Db     639566 TACATTACACCAGTCCCCGGTGGCGTTGGGCCAATGACCGTTGCAATGTTGATGAGCAAT 639507

Qy        281 ThrLeu 282
              ||||||
Db     639506 ACGCTT 639501

RESULT 9
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;   GENERAL INFORMATION:
;        APPLICANT: Robert D. Fleischmann
;                   Mark D. Adams
;                   Owen White
;                   Hamilton O. Smith
;                   J. Craig Venter
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville,
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/643,990A
;             FILING DATE: 23-Aug-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/487,429
;             FILING DATE: 1995-06-07
;             APPLICATION NUMBER: 08/426,787
;             FILING DATE: 1995-04-21
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kenley K. Hoover
;             REGISTRATION NUMBER: 40,302
;             REFERENCE/DOCKET NUMBER: PB186P1C1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-610-5790
;             TELEFAX: 310-309-8439
```

```
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.:              2.68e-70      Length:         1830121
Score:                  681.50        Matches:        144
Percent Similarity:     66.67%        Conservative:   44
Best Local Similarity:  51.06%        Mismatches:     85
Query Match:            45.86%        Indels:         9
DB:                     4             Gaps:           2

US-10-723-061-18 (1-292) x US-09-643-990A-1 (1-1830121)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              ||||||||||||  |||      ::::::|||||||||:::      :::|||   :::
Db     640319 GCAAAAATCATCTCTGGTACTGAACTTTCAAAACAAATTAAAGCTAATTTAGCAGATAAA 640260

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              |||        ::                |||||||||||||||||||||||||
Db     640259 ATAACCCATTATATTGAACAAGGAAAACGCGCTCCAGGACTTGCTGTAATCCTCGTTGGT 640200

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              :::       ||||||  ||||||  ||||||:::|||:::||||||  ||||||
Db     640199 GCTGATCCTGCTTCGCAAATTTATGTGGGCAATAAACGTAAAAGTTGCGAGGAAGTCGGT 640140

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||  |||         :::||||||||  ::::::::::|||:::|||
Db     640139 ATTCTCTCTAAATCTTATGACCTCCCTGAAACAACCACGCAAAATGAATTATTAGCGATT 640080

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              |||         |||:::|||  ::::::  ||||||||||||||||||||||||||
Db     640079 ATTGACCAACTTAACGCAGATAAAAATATTGACGGTATTTTGGTGCAACTCCCCTTGCCA 640020

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                :::|||  |||  |||:::      :::      :::|||||||||||||||
Db     640019 AAACAAATTAATGCGGAAGCTATTATTGAACGTATCGATCCTAAAAAAGATGTAGATGGC 639960

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||  |||:::|||||||||          |||   |||
Db     639959 TTTCATCCTTATAATGTTGGGCGGTTATGCCAA------CGTATTCCAACTTTACGTGCT 639906

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||||||||   |||   |||:::||||||          |||:::::::::   ||||||
Db     639905 TGTACACCTTATGGTGTGATGAAACTATTAGAAACAACGGGTATTGATTTACACGGCAAG 639846

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||::::::|||   ||||||||||||||||  |||   :::|||   |||
Db     639845 CACGCAGTAATTGTAGGCGCATCCAATATTGTCGGCCGCCCAATGTCTTTAGAATTATTA 639786

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
                |||   |||||||||::::::   |||       ||||||      |||
Db     639785 TTAGCAGGGGCAACCGTCACAGTAACACATAGATTCACTAAAAATTTGGAAAACCACGTT 639726

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              ||||||||||||:::::::   |||||||||       ||||||:::   |||   |||
Db     639725 CGCCAAGCAGATATACTTGTCGTTGCGGTAGGTAAACCAAATTTAATTTCTGGCGATTGG 639666

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||     |||   :::|||||||||||||||||||   ||||||
Db     639665 ATCAAAGAAAGTGCGGTAGTAATTGATGTTGGAATTAACCGAGTAGATGGA--------- 639615

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                        :::|||||||||||||||   :::::::::|||::::::   |||
Db     639614 ------------AAATTAGTAGGCGATGTAGAGTTTGATAAAGCAGCCGAAAAAGCGGCT 639567

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||:::|||||||||:::||||||

Db     639566 TACATTACACCAGTCCCCGGTGGCGTTGGGCCAATGACCGTTGCAATGTTGATGAGCAAT 639507

Qy        281 ThrLeu 282
              ||||||
Db     639506 ACGCTT 639501

RESULT 10
US-09-134-001C-564
; Sequence 564, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 564
;   LENGTH: 873
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-564

Alignment Scores:
Pred. No.:              1.92e-75      Length:         873
Score:                  679.50        Matches:        139
Percent Similarity:     69.23%        Conservative:   59
Best Local Similarity:  48.60%        Mismatches:     79
Query Match:            45.73%        Indels:         9
DB:                     4             Gaps:           3

US-10-723-061-18 (1-292) x US-09-134-001C-564 (1-873)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||:::|||||||||   :::||||||:::   |||:::   :::      :::
Db         19 GCTAAAATTTTAGATGGTAAACAAATAGCAAAAGAATATAGACAAAGACTTAAAAATCAA 78

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::       :::|||:::      |||       |||   |||:::|||||||||||||||
Db         79 GTCAATGACTTAAAAGAA---TATGGTTTTACTCCAAAATTATCAGTCATATTAGTTGGT 135

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              :::          ||||||:::||||||::::::||||||||||||   |||   :::|||
Db        136 AATGATGGTGCAAGCCAAAGTTATGTAAAATCAAAGAAAAAAGCAGCAGAAAAAATTGGA 195

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              :::   |||       ::::::|||   |||   :::|||||||:::   ||||||
Db        196 ATGATATCAGAAATTATTCACTTAGACGAATCCACATCAGAAGAAGTAGTTTTAAGCGAA 255

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::         |||:::|||   :::|||   ||||||||||||||:::||||||||
Db        256 CTTAATCGATTAAACAATGACGATACTGTTAGTGGTATTTTAGTTCAAGTACCATTACCT 315

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                  ::::::::::::   :::|||   |||::::::   ||||||||||||||||||
Db        316 AAACAGGTAAGCGAACAAAAAGTACTTGAAGCTATCAATCCGGAAAAAGATGTAGATGGT 375

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||:::||||||||:::|||   :::         :::       |||||||||
Db        376 TTCCATCCAATTAATATTGGAAAATTGTATATTGAT------GAGCAAACCTTTGTACCA 429

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||||||||   |||   ||||||:::|||           :::   ::::::||||||
Db        430 TGTACACCCCTGGGTATTATGGAAATTTTAAAACACGCAGATATTAATTTAGAAGGAAAA 489

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||:::||||||||   |||   :::   ||||||
Db        490 AATGCAGTTGTTATTGGTCGTAGTCATATTGTTGGCCAACCTGTTTCAAAGTTATTGTTG 549
```

```
Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::|||||||||||||||||:::|||:::||||||   ||||||
Db        550 CAAGCTAACGCAACTGTTACGATTTTACATTCACGTACAAAAAATATGAATGCACACTTA 609

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              :::||||||||:::::::::::|||||||||        ||||||
Db        610 AAACAAGCAGATGTGATTGTTAGCGCAGTAGGACAACCTGGTTTAGTTACTAAAGAAAAT 669

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              :::|||   ||||||   ||||||||||||||||      |||   ::::::
Db        670 GTCAAAAAAGGCGCAGTAATTATAGATGTTGGTAATACACCTGATGAAAATGGA------ 723

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                        :::|||   |||||||||   |||:::|||      :::||||||
Db        724 ------------AAATTAAAAGGAGATGTAGCATATGATGAAGTAAAAGAAATAGCTAGT 771

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||:::||||||   |||:::|||:::|||
Db        772 GCAATTACTCCAGTACCAGGTGGTGTTGGTCCTTTAACTATTACAATGGTACTTAATAAT 831

Qy        281 ThrLeuGluSerAlaLys 286
              ||||||   :::   |||
Db        832 ACACTTCTTGCAGAGAAA 849

RESULT 11
US-09-252-991A-13712
; Sequence 13712, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 13712
;   LENGTH: 957
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13712

Alignment Scores:
Pred. No.:              4.59e-75        Length:         957
Score:                  677.00          Matches:        144
Percent Similarity:     64.36%          Conservative:   42
Best Local Similarity:  49.83%          Mismatches:     95
Query Match:            45.56%          Indels:         8
DB:                     4               Gaps:           2

US-10-723-061-18 (1-292) x US-09-252-991A-13712 (1-957)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||::::::||||||||||||   :::|||      :::|||::::::||||||
Db        109 GCACAACTGATCGACGGCAAAGCGATCGCCGCCAACCTTCGCCAGCAGATAGCCCAACGC 168

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::   :::   :::                ||||||||||||||||||||||||||||
Db        169 GTGACCGAGCGCCGCCAGCAAGGCCTGCGCGTTCCCGGCCTGGCGGTGATCCTGGTCGGC 228

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              :::      ||||||   ||||||   :::|||:::|||   ||||||   ||||||
Db        229 ACCGATCCGGCCTCTCAGGTCTATGTGGCGCACAAGCGCAAGGACTGCGAGGAAGTCGGC 288

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
                 |||      :::||||||   ::::::|||:::|||:::::::|||
Db        289 TTTCTCTCCCAGGCCTACGATCTTCCCGCCGAAACCAGCCAGGACGACCTGCTGGCCCTG 348

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              |||         |||   ||||||::::::   ||||||||||||||||||||||||
Db        349 ATCGACCGCCTGAACGACGACCCCGCCATCGACGGCATCCTGGTCCAGCTACCCCTGCCC 408

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                 |||::::::      :::|||   :::      :::|||||||||||||||
Db        409 GCCCACCTGGACGCCTCCCTGCTGCTGGAGCGCATCCACCCGGACAAGGACGTGGACGGT 468

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||   ||||||||||||||||||          |||   |||          |||
Db        469 TTCCATCCCTACAACATCGGCCGCCTGGCCCAG------CGCATGCCCCTCCTGCGCCCC 522

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||   |||   ||||||          |||   ::::::   |||
Db        523 TGCACCCCGAAAGGCATCATGACCCTGCTCGCCAGCACCGGCGCCGACCTGTACGGCATG 582

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
                 |||||||||:::|||   |||||||||||||||   |||   ||||||   |||
Db        583 GATGCGGTCGTGGTCGGCGCCTCGAACATCGTCGGCCGGCCCATGGCTCTGGAGTTGCTG 642

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
                       ||||||::::::   |||       |||:::          :::
Db        643 CTGGGTGGCTGCACCGTCACCGTGACCCACCGCTTCACCCGCGACCTGGCCGACCATGTG 702

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              :::||||||:::|||:::   |||   |||            ||||||:::|||   |||
Db        703 TCGCGCGCCGACCTGGTGGTGGTCGCTGCCGGCAAGCCGGGACTGGTCAAGGGCGAGTGG 762

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||   ||||||   :::||||||||||||||||||          |||
Db        763 ATCAAGGAAGGCGCCATCGTCATCGACGTCGGCATCAACCGCCAGGCCGACGGC------ 816

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                        ||||||||||||||||||   ||||||   |||::::::   |||
Db        817 ------------CGCCTGGTCGGCGACGTGGAATACGAAGTGGCAGCGCAACGCGCCAGC 864

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||   |||   ||||||   |||
Db        865 TGGATCACCCCGGTGCCGGGCGGCGTCGGGCCGATGACCCGCGCCTGCCTGCTGGAAAAT 924

Qy        281 ThrLeuGluSerAlaLysArgIleHis 289
              ||||||   :::|||:::   :::|||
Db        925 ACCCTGCACGCCGCCGAACACCTGCAC 951

RESULT 12
US-09-252-991A-13926
; Sequence 13926, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 13926
;   LENGTH: 1188
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13926

Alignment Scores:
Pred. No.:              6.53e-75        Length:         1188
Score:                  677.00          Matches:        144
Percent Similarity:     64.36%          Conservative:   42
Best Local Similarity:  49.83%          Mismatches:     95
Query Match:            45.56%          Indels:         8
DB:                     4               Gaps:           2
```

```
US-10-723-061-18 (1-292) x US-09-252-991A-13926 (1-1188)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||::::::|||||||||||   :::|||      :::|||::::::||||||
Db        260 GCACAACTGATCGACGGCAAAGCGATCGCCGCCAACCTTCGCCAGCAGATAGCCCAACGC 319

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::   :::   :::                |||||||||||||||||||||||||||||
Db        320 GTGACCGAGCGCCGCCAGCAAGGCCTGCGCGTTCCCGGCCTGGCGGTGATCCTGGTCGGC 379

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              :::      ||||||   ||||||   :::|||:::|||   ||||||   ||||||
Db        380 ACCGATCCGGCCTCTCAGGTCTATGTGGCGCACAAGCGCAAGGACTGCGAGGAAGTCGGC 439

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
                 |||         :::||||||   ::::::|||:::|||::::::|||
Db        440 TTTCTCTCCCAGGCCTACGATCTTCCCGCCGAAACCAGCCAGGACGACCTGCTGGCCCTG 499

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              |||         |||   ||||||::::::   ||||||||||||||||||||||||||
Db        500 ATCGACCGCCTGAACGACGACCCCGCCATCGACGGCATCCTGGTCCAGCTACCCCTGCCC 559

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                 |||::::::         :::|||      :::      :::|||||||||||||||
Db        560 GCCCACCTGGACGCCTCCCTGCTGCTGGAGCGCATCCACCCGGACAAGGACGTGGACGGT 619

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||   ||||||||||||||||||          |||   |||          |||
Db        620 TTCCATCCCTACAACATCGGCCGCCTGGCCCAG------CGCATGCCCCTCCTGCGCCCC 673

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||   |||   ||||||         |||   ::::::   |||
Db        674 TGCACCCCGAAAGGCATCATGACCCTGCTCGCCAGCACCGGCGCCGACCTGTACGGCATG 733

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
                 |||||||||:::|||   ||||||||||||||   |||   ||||||   |||
Db        734 GATGCGGTCGTGGTCGGCGCCTCGAACATCGTCGGCCGGCCCATGGCTCTGGAGTTGCTG 793

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
                       ||||||::::::   |||       |||:::          :::
Db        794 CTGGGTGGCTGCACCGTCACCGTGACCCACCGCTTCACCCGCGACCTGGCCGACCATGTG 853

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
                 :::||||||:::|||:::   |||   |||          ||||||:::|||   |||
Db        854 TCGCGCGCCGACCTGGTGGTGGTCGCTGCCGGCAAGCCGGGACTGGTCAAGGGCGAGTGG 913

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||   ||||||   :::||||||||||||||||||          |||
Db        914 ATCAAGGAAGGCGCCATCGTCATCGACGTCGGCATCAACCGCCAGGCCGACGGC------ 967

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                       |||||||||||||||||   ||||||   |||::::::   |||
Db        968 -----------CGCCTGGTCGGCGACGTGGAATACGAAGTGGCAGCGCAACGCGCCAGC 1015

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              |||||||||||||||||||||||||||||||||||   |||   ||||||   |||
Db       1016 TGGATCACCCCGGTGCCGGGCGGCGTCGGGCCGATGACCCGCGCCTGCCTGCTGGAAAAT 1075

Qy        281 ThrLeuGluSerAlaLysArgIleHis 289
              ||||||   :::|||:::   :::|||
Db       1076 ACCCTGCACGCCGCCGAACACCTGCAC 1102

RESULT 13
US-08-956-171E-85
; Sequence 85, Application US/08956171E
; Patent No. 6593114
;   GENERAL INFORMATION:
;        APPLICANT: Charles Kunsch
;                   Gil H. Choi
;                   Patrick S. Dillon
;                   Craig A. Rosen
;                   Steven C. Barash
;                   Michael R. Fannon
;        TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;        NUMBER OF SEQUENCES: 5256
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/956,171E
;             FILING DATE: 20-Oct-1997
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/009,861
;             FILING DATE: January 5, 1996
;             APPLICATION NUMBER: 08/781,986
;             FILING DATE: January 3, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark J. Hyman
;             REGISTRATION NUMBER: 46,789
;             REFERENCE/DOCKET NUMBER: PB248P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (240) 314-1224
;             TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 85:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 9062 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-08-956-171E-85

Alignment Scores:
Pred. No.:                1.16e-72        Length:          9062
Score:                    670.50          Matches:         139
Percent Similarity:       68.88%          Conservative:    58
Best Local Similarity:    48.60%          Mismatches:      80
Query Match:              45.12%          Indels:          9
DB:                       4               Gaps:            3

US-10-723-061-18 (1-292) x US-08-956-171E-85 (1-9062)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||:::|||||||||   :::||||||      |||:::   :::      :::
Db       6676 GCTAAAATTTTAGATGGTAAACAAATTGCCAAAGACTACAGACAGGGGTTACAAGATCAA 6735

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::      :::|||:::      |||      |||   |||:::|||||||||||||||
Db       6736 GTTGAAGCGCTAAAAGAAAAG---GGTTTTACACCTAAATTATCCGTTATATTAGTTGGT 6792

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              :::         ||||||:::|||||||||:::||||||||||||   |||   :::|||
Db       6793 AATGATGGCGCTAGTCAAAGTTATGTTAGATCAAAAAAGAAAGCAGCTGAAAAAATTGGT 6852

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              :::   |||      |||:::|||   |||   ::::::|||:::|||||||||
Db       6853 ATGATTTCAGAAATCGTACATTTGGAAGAAACAGCTACTGAAGAAGAAGTATTAAACGAA 6912

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::         |||:::|||   ||||||   ||||||||||||||:::|||||||||
Db       6913 CTAAATAGACTAAATAATGATGATTCTGTAAGTGGTATTTTGGTACAAGTACCATTACCA 6972

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                  ::::::::::::   ||||||   |||::::::   |||||||||||||||||
Db       6973 AAACAAGTTAGCGAACAGAAAATATTAGAAGCAATCAATCCTGAAAAAGATGTGGACGGT 7032
```

```
Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
Db       7033 TTTCATCCAATAAATATAGGGAAATTATATATCGAT------GAACAAACTTTTGTACCT 7086

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db       7087 TGCACACCGCTCGGCATCATGGAAATATTAAAACATGCTGATATTGATTTAGAAGGTAAA 7146

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
Db       7147 AATGCAGTTGTAATTGGACGAAGTCATATTGTCGGACAACCAGTTTCTAAGTTACTACTT 7206

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db       7207 CAAAAAAATGCATCAGTAACAATCTTACATTCTCGTTCAAAAGATATGGCATCATATTTA 7266

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db       7267 AAAGATGCTGATGTCATTGTCAGTGCAGTTGGTAAGCCTGGTTTAGTAACAAAAGATGTG 7326

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db       7327 GTCAAAGAAGGAGCAGTAATTATCGATGTTGGCAATACGCCAGATGAAAATGGC------ 7380

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db       7381 ------------AAATTAAAAGGTGACGTTGATTATGATGCGGTTAAAGAAATTGCTGGA 7428

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db       7429 GCTATTACACCAGTTCCTGGTGGCGTTGGTCCATTAACAATTACTATGGTATTAAATAAT 7488

Qy        281 ThrLeuGluSerAlaLys 286
Db       7489 ACTTTGCTTGCAGAAAAA 7506
```

RESULT 14
US-09-134-000C-460
; Sequence 460, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 460
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-460

Alignment Scores:

Pred. No.:	1.83e-73	Length:	840
Score:	663.50	Matches:	135
Percent Similarity:	68.52%	Conservative:	50
Best Local Similarity:	50.00%	Mismatches:	76
Query Match:	44.65%	Indels:	9
DB:	4	Gaps:	5

US-10-723-061-18 (1-292) x US-09-134-000C-460 (1-840)

```
Qy          3 IleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIleAla 22
Db         34 GTAATTAACGGTCGTGAATTAGCAGATCAAATGCAAGCAGAGATTCAAAAAGATGTAGAA 93

Qy         23 LysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySerArg 42
Db         94 AAGATGACACAACAA---GGCATCCAACCAGGATTAGTTGTTTTATTAGTTGGGGAAAAT 150

Qy         43 LysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIleLys 62
Db        151 CCTGCCAGCCAAACGTATGTGAGAAATAAAGAACGTGCAGCAGCCAAAATTGGCATTCTG 210

Qy         63 SerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAla 82
Db        211 TCAAAGGTCGAAAAACTGCCAGAAACTATTTCAGAAGAAGAATTATTGGCTGAGATTGAC 270

Qy         83 ThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHis 102
Db        271 AAATATAATCAAGATTCACGCTTTCATGGCATTCTTGTACAACTACCTTTGCCAAAACAT 330

Qy        103 MetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHis 122
Db        331 ATTGATGAAGAAAAGATTTTATTAGCGATTGATCCCAAAAAAGATGTAGATGGGTTCCAT 390

Qy        123 ProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThr 142
Db        391 CCAATGAATTTAGGCCGTTTGTTTGTA---GGCAAA---CCTGAAATGATTCCTTGCACG 444

Qy        143 ProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAla 162
Db        445 CCGTATGGAATTATGAAGATGTTTGAAGCTTATGATATTGATTTAACAGGTAAACGTGCG 504

Qy        163 ValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLysAla 182
Db        505 GTGGTTATTGGTCGAAGCAATATTGTTGGCAAGCCAATGGCACAATTATTGTTAATGAAA 564

Qy        183 AsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGln 202
Db        565 AATGCGACCGTGACGATTGCCCATTCTAAAACTGAACATTTAGCAGAAGTTGCAAAAGAA 624

Qy        203 AlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLys 222
Db        625 GCGGATATTTTAGTTGTAGCAATTGGACGCGGGCATTTTGTCACCAAAGAGTTCGTGAAA 684

Qy        223 ProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerProArg 242
Db        685 CCAGGTGCGGTAGTGATTGATGTGGGCATGAAT--------------CGCAATCAAGAA 729

Qy        243 GlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIle 262
Db        730 GGA---AAACTGATTGGTGATGTAGCATTCGATGAAGTTTCTGAAATTGCAAGCTATATC 786

Qy        263 ThrProValProGlyGlyValGlyProMet 272
Db        787 ACGCCGGTTCCCAAAGGTGTTGGCCCAATG 816
```

RESULT 15
US-09-543-681A-2906
; Sequence 2906, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILI
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2906
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2906

Alignment Scores:

Pred. No.:	4.01e-73	Length:	873

Score:	661.00	Matches:	140
Percent Similarity:	64.48%	Conservative:	47
Best Local Similarity:	48.28%	Mismatches:	93
Query Match:	44.48%	Indels:	10
DB:	4	Gaps:	4

US-10-723-061-18 (1-292) x US-09-543-681A-2906 (1-873)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||:::|||||||||||||||||   :::|||:::   ||||||   |||:::|||   |||
Db          7 GCAAGAATTATAGATGGGAAAACGATTGCGCCGACCATCAGAAGTGAAGTGGCA---GAA 63

Qy         21 IleAlaLysMetLysAspAlaIleGlyVal---ValProGlyLeuAlaValIleLeuVal 39
                  |||   :::      ||||||          ||||||||||||||||||||||||
Db         64 AAAGTAAAACAACGTATTAAGATAGGAAAACGTGCACCGGGTTTAGCGGTTATTTTAGTC 123

Qy         40 GlySerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaVal 59
              |||             ||||||   ||||||   :::|||:::|||||||||:::   |||
Db        124 GGTGATAACCCTGCATCGCAAATCTATGTTGCTAGTAAACGTAAAGCTTGTGATGAGGTA 183

Qy         60 GlyIleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLys 79
              |||       |||          :::|||||||:::   :::||||||   ::::::|||
Db        184 GGATTTATCTCTCGCTCTTACGATCTACCGGATACTACAAGTGAAGCAGATTTACTCAAT 243

Qy         80 HisIleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeu 99
                 |||   |||   |||   |||   ::::::   ||||||||||||||||||||||
Db        244 CTCATTGATACTCTAAATGAAGATAATACTATTGATGGTATTTTAGTGCAGCTCCCTTTA 303

Qy        100 ProHisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAsp 119
              |||       :::::::::      :::|||      :::      :::|||||||||||
Db        304 CCTGCGGGAATTGATAACGTCAAAGTATTAGAACGTATTCATCCTGATAAAGATGTTGAT 363

Qy        120 GlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheVal 139
              ||||||||||||   ||||||||||||||            |||   |||
Db        364 GGCTTCCATCCTTACAATATTGGTCGGTTGTGCCAA------CGTGCACCTAAGCTACGT 417

Qy        140 ProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGly 159
              |||||||||||:::|||   :::   ||||||   |||      :::   :::   |||
Db        418 CCTTGCACCCCTCGTGGTATTGTCACACTCTTAGAGCGTTGCAATATTCCAATGAATGGC 477

Qy        160 LysArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeu 179
                  ||||||:::||||||   ||||||||||||||   |||   :::|||   |||
Db        478 TTAAATGCGGTAATTATTGGTGCTTCAAATATTGTTGGTCGCCCAATGAGTTTAGAGCTA 537

Qy        180 GlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIle 199
                 |||      |||   ::::::   |||      ||||||
Db        538 CTGCTTGCAGGCTGTACAACGACTGTCACTCATCGTTTTACTAAAGACTTACGTTTTCAT 597

Qy        200 ThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySer 219
                    ||||||:::|||:::   |||||||||      |||   :::   |||
Db        598 GTTGAACATGCTGATTTGGTGGTTGTGGCAGTTGGTAAACCTAACTTTATTCCTGGTGAA 657

Qy        220 TrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGlu 239
              ||||||||||||||||||   :::|||||||||||||||||   :::::::::
Db        658 TGGATAAAACCTGGCGCTATTGTCATTGATGTCGGTATTAATCGCCTTGAAAACGGT--- 714

Qy        240 SerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAla 259
                        ::::::||||||||||||||   :::   :::||||||:::   |||
Db        715 ---------------AAAGTGGTTGGTGATGTTGATTTCGCACAAGCCTCTCAACGCGCG 759

Qy        260 GlyLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSer 279
              |||   |||:::||||||||||||||||||||||||||||||:::|||   |||:::
Db        760 GGGTGGATCTCCCCTGTTCCGGGTGGTGTAGGTCCAATGACTGTAGCGACTTTGATTCAA 819

Qy        280 AsnThrLeuGluSerAlaLysArgIleHis 289
              |||||||||::::::   :::      |||
Db        820 AATACGTTGCAAGCCTGCGAAGAATATCAC 849
```

Search completed: July 24, 2004, 10:04:20
Job time : 1148 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 24, 2004, 08:50:33 ; Search time 575 Seconds
(without alignments)
2482.406 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3216467 seqs, 2444149694 residues

Total number of hits satisfying chosen parameters: 6432934

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10723061/runat_21072004_090403_7226/app_query.fasta_1.455
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10723061_@CGN_1_1_912_@runat_21072004_090403_7226
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1238	83.3	1083	9	US-09-938-842A-541	Sequence 541, App
2	1238	83.3	1083	11	US-09-938-842A-541	Sequence 541, App
3	1234	83.0	1632	13	US-10-424-599-91222	Sequence 91222, A
4	1030	69.3	926	9	US-09-903-814-19	Sequence 19, Appl
5	1030	69.3	926	17	US-10-723-061-19	Sequence 19, Appl
6	1025	69.0	1066	13	US-10-425-114-33029	Sequence 33029, A
7	1025	69.0	1415	13	US-10-425-114-23607	Sequence 23607, A
8	1019	68.6	1255	13	US-10-425-114-11296	Sequence 11296, A
9	1019	68.6	1798	13	US-10-424-599-114288	Sequence 114288,
10	1016	68.4	1076	9	US-09-903-814-15	Sequence 15, Appl
11	1016	68.4	1076	17	US-10-723-061-15	Sequence 15, Appl
12	1012	68.1	1356	17	US-10-437-963-74482	Sequence 74482, A
13	1010	68.0	1286	13	US-10-424-599-125693	Sequence 125693,
14	994	66.9	1409	13	US-10-425-114-9450	Sequence 9450, Ap
15	988	66.5	1413	17	US-10-437-963-23018	Sequence 23018, A
16	965	64.9	1561	13	US-10-425-114-28649	Sequence 28649, A
17	919	61.8	2442	17	US-10-437-963-52595	Sequence 52595, A
18	801	53.9	837	13	US-10-425-114-2899	Sequence 2899, Ap
19	775	52.2	837	16	US-10-369-493-33838	Sequence 33838, A
c 20	763.5	51.4	495269	16	US-10-398-221-8	Sequence 8, Appli
21	763.5	51.4	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
22	751.5	50.6	5635	16	US-10-398-221-3823	Sequence 3823, Ap
23	740.5	49.8	875	9	US-09-903-814-21	Sequence 21, Appl
24	740.5	49.8	875	17	US-10-723-061-21	Sequence 21, Appl
25	735	49.5	3112	9	US-09-880-107-2103	Sequence 2103, Ap
26	735	49.5	3112	12	US-09-968-007A-367	Sequence 367, App
27	735	49.5	3112	15	US-10-171-581-40	Sequence 40, Appl
28	730	49.1	1035	16	US-10-369-493-40551	Sequence 40551, A
29	724	48.7	858	10	US-09-769-787-239	Sequence 239, App
30	724	48.7	13440	13	US-10-158-844-128	Sequence 128, App
31	722.5	48.6	891	16	US-10-369-493-45509	Sequence 45509, A
32	716	48.2	840	16	US-10-369-493-41046	Sequence 41046, A
33	715	48.1	900	16	US-10-369-493-35899	Sequence 35899, A
34	711	47.8	855	16	US-10-369-493-43170	Sequence 43170, A
35	710.5	47.8	840	16	US-10-369-493-42987	Sequence 42987, A
36	707.5	47.6	885	16	US-10-369-493-42077	Sequence 42077, A
37	704.5	47.4	7528	9	US-09-070-927A-55	Sequence 55, Appl
38	694.5	46.7	834	16	US-10-369-493-41608	Sequence 41608, A
39	690.5	46.5	901	16	US-10-369-493-31770	Sequence 31770, A
40	687.5	46.3	852	16	US-10-369-493-34559	Sequence 34559, A
41	687.5	46.3	855	9	US-09-974-300-998	Sequence 998, App
42	684.5	46.1	840	16	US-10-369-493-40237	Sequence 40237, A
43	683.5	46.0	840	16	US-10-369-493-30840	Sequence 30840, A
44	683.5	46.0	852	16	US-10-369-493-46889	Sequence 46889, A
c 45	681.5	45.9	1830121	15	US-10-329-960-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-938-842A-541
; Sequence 541, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 541
; LENGTH: 1083

```
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-541

Alignment Scores:
Pred. No.:              1.88e-146       Length:         1083
Score:                  1238.00         Matches:        231
Percent Similarity:     91.44%          Conservative:   36
Best Local Similarity:  79.11%          Mismatches:     25
Query Match:            83.31%          Indels:         0
DB:                     9               Gaps:           0

US-10-723-061-18 (1-292) x US-09-938-842A-541 (1-1083)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::|||||||||||||   |||||||||:::|||||:::|||||   :::|||
Db        205 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 264

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::|||||:::::|||||||||:::||||||||||||||||||||||||||||
Db        265 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 324

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
                 ||||||||||||   |||||||||||||||||||||||||||:::::|||||||
Db        325 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 384

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              ||||||||||:::||||||   |||   |||||||||||||||||:::||||||||||||
Db        385 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 444

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::   ||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        445 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 504

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              ||||||::::::::::||||||||||||||||||||||||||||||||||||||||||||
Db        505 TCGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATAGAGAAAGATGTTGACGGA 564

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||||||||||||||:::|||||:::|||   |||||||||
Db        565 TTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGAGAACCCTTATTCGTTCCT 624

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||:::||||||||||||||||   :::||||||||||||||||||
Db        625 TGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAACATTGAAATCAAAGGAAAG 684

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCAGCTGCTCTTTTACTGCAG 744

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::||||||||||||||||:::||||||   ||||||   ||||||||||||||||
Db        745 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 804

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||:::|||:::|||||||||      |||:::||||||||||||||||
Db        805 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 864

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||   :::||||||||||||||||||||||||||:::|||||||   :::
Db        865 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGAGGATCCAAGTGCT 924

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||||:::|||||||||||||||||||||:::|||
Db        925 GCGCGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCA 984

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 GCCATCACACCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1044

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||   ||||||||||||||||||   |||:::
Db       1045 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1080
```

```
RESULT 2
US-09-938-842A-541
; Sequence 541, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 541
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-541

Alignment Scores:
Pred. No.:              1.88e-146       Length:         1083
Score:                  1238.00         Matches:        231
Percent Similarity:     91.44%          Conservative:   36
Best Local Similarity:  79.11%          Mismatches:     25
Query Match:            83.31%          Indels:         0
DB:                     11              Gaps:           0

US-10-723-061-18 (1-292) x US-09-938-842A-541 (1-1083)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::|||||||||||||   |||||||||:::|||||:::|||||   :::|||
Db        205 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 264

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::|||||:::::|||||||||:::||||||||||||||||||||||||||||
Db        265 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 324

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
                 ||||||||||||   |||||||||||||||||||||||||||:::::|||||||
Db        325 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 384

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              ||||||||||:::||||||   |||   |||||||||||||||||:::||||||||||||
Db        385 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 444

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::   ||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        445 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 504

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              ||||||::::::::::||||||||||||||||||||||||||||||||||||||||||||
Db        505 TCGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATAGAGAAAGATGTTGACGGA 564

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||||||||||||||:::|||||:::|||   |||||||||
Db        565 TTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGAGAACCCTTATTCGTTCCT 624

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||:::||||||||||||||||   :::||||||||||||||||||
Db        625 TGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAACATTGAAATCAAAGGAAAG 684

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCAGCTGCTCTTTTACTGCAG 744
```

```
Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::|||||||||||||||||:::||||||   ||||||   ||||||||||||||||
Db        745 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 804

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||:::|||:::|||||||||      |||:::|||||||||||||||
Db        805 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 864

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||   :::|||||||||||||||||||||||||:::||||||   :::
Db        865 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGAGGATCCAAGTGCT 924

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||:::|||||||||||||||||||||:::|||
Db        925 GCGCGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCA 984

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 GCCATCACACCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1044

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||   ||||||||||||||||||   |||:::
Db       1045 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1080
```

RESULT 3
US-10-424-599-91222
; Sequence 91222, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91222
; LENGTH: 1632
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53384C.1
US-10-424-599-91222

Alignment Scores:

Pred. No.:	1.13e-145	Length:	1632
Score:	1234.00	Matches:	233
Percent Similarity:	89.73%	Conservative:	29
Best Local Similarity:	79.79%	Mismatches:	30
Query Match:	83.04%	Indels:	0
DB:	13	Gaps:	0

US-10-723-061-18 (1-292) x US-10-424-599-91222 (1-1632)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              ||||||:::|||||||||||||   |||||||||||||||||||:::||||||      |||
Db        309 GCTAAGGTGATTGATGGAAAATCCGTTGCCAAGCAAATCAGAGATGAGATAACGGCTGAG 368

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              ::::::::::|||::::::::::|||||||||:::|||||||||||||||||||||||||
Db        369 GTTTCCAGGATGAGAGAATCCATTGGTGTGATTCCTGGGTTGGCTGTAATTCTTGTTGGG 428

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||   |||||||||||||||||||||||||||||||:::||||||
Db        429 GATAGGAAAGACTCAGCAACTTATGTGCGTAACAAGAAGAAGGCTTGTGAATCTGTTGGA 488

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||   |||   |||   |||||||||||||||||:::|||:::||||||||||   :::
Db        489 ATCAATTCTTTGGAAGCAAATCTGCCAGAGGATTCCACAGAAGAAGAAGTTTTGAACTAT 548
```

```
Qy          81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ||||||   :::|||   |||||||||||||||||||||||||||||||||||||||||||
Db        549 ATTGCAGGCTACAATGATGATCCTTCAGTTCATGGCATCCTTGTTCAGTTACCCTTACCT 608

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||||||||::::::|||||||||||||||||||   ||||||||||||||||||||||
Db        609 TCGCATATGAATGAGCAGAACATCTTGAACGCTGTCAGGATTGAGAAGGATGTAGATGGT 668

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||||||||||||||:::||||||:::|||   |||||||||
Db        669 TTTCATCCGTTAAATATTGGTCGTCTTGCTATGCGTGGAAGAGAACCTCTGTTTGTTCCT 728

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||:::||||||||||||||||||   |||   ||||||||||||
Db        729 TGTACACCAAAGGGATGCATAGAGCTACTGCACAGATACAATGTTTCTATTAAAGGAAAG 788

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        789 AGGGCTGTTGTGATTGGTCGGAGCAATATTGTTGGAATGCCAGCTGCTCTCTTGCTGCAA 848

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::|||||||||||||||||||||||||   |||      ||||||||||||||
Db        849 AGGGAAGATGCTACTGTCAGTATTGTCCATTCTAGAACCAGTAACCCCGAAGAGATCATA 908

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              ||||||||||||||:::|||||||||||||||   ||||||:::||||||||||||||||
Db        909 AGACAGGCAGATATTATCATTGCTGCCGTTGGGCAAGCAAACATGGTGAGGGGAAGCTGG 968

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||   |||||||||||||||||||||||||||:::||||||   |||
Db        969 ATAAAGCCTGGTGCAGTCATTATTGACGTCGGAATCAACCCAGTAGAGGATCCAAATAGT 1028

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||:::|||||||||||||||||||||||||||||||||   :::|||||||
Db       1029 CCCCGAGGTTACAAACTGGTGGGAGATGTTTGTTATGAAGAAGCCATAAGAATTGCCTCT 1088

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1089 GCTGTTACACCAGTTCCTGGAGGAGTTGGTCCAATGACCATAGCAATGCTTCTCCAAAAT 1148

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||   ||||||||||||:::|||   |||:::
Db       1149 ACACTCATTTCTGCAAAGAGGGTGCACCATTTTGAA 1184
```

RESULT 4
US-09-903-814-19
; Sequence 19, Application US/09903814
; Patent No. US20020102689A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB-1179-A
; CURRENT APPLICATION NUMBER: US/09/903,814
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/092,869
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 926
; TYPE: DNA
; ORGANISM: Glycine max
US-09-903-814-19

Alignment Scores:

Pred. No.:	3.35e-120	Length:	926
Score:	1030.00	Matches:	194
Percent Similarity:	89.63%	Conservative:	22

```
Best Local Similarity:  80.50%          Mismatches:     25
Query Match:            69.31%          Indels:         0
DB:                     9               Gaps:           0

US-10-723-061-18 (1-292) x US-09-903-814-19 (1-926)

Qy         52 LysLysLysAlaCysGluAlaValGlyIleLysSerTyrGluValAsnLeuProGluAsp 71
              :::|||||||||||||||:::|||||||||   |||   |||   ||||||||||||:::
Db          6 GAGAAGAAGGCTTGTGAATCTGTTGGAATCAATTCTTTGGAAGCAAATCTGCCTGAGAAT 65

Qy         72 SerSerGluAspGluValLeuLysHisIleAlaThrPheAsnSerAspProSerValHis 91
              |||:::|||:::||||||||||   :::||||||   :::|||   ||||||||||||||
Db         66 TCCACAGAAGAAGAAGTTTTGAACTATATTGCAGGCTACAATGATGATCCTTCAGTTCAT 125

Qy         92 GlyIleLeuValGlnLeuProLeuProHisHisMetAsnAspGluAsnIleLeuAsnAla 111
              |||||||||||||||||||||||||   |||||||||::::::||||||||||||||||
Db        126 GGCATCCTCGTTCAGTTACCCTTACCTTCGCATATGAATGAGCAGAACATCTTGAACGCT 185

Qy        112 ValSerIleGluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMet 131
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 GTCAGGATTGAGAAGGATGTAGATGGTTTTCATCCGTTAAATATTGGTCGTCTTGCCATG 245

Qy        132 GlnGlyArgAspProPhePheValProCysThrProLysGlyCysMetGluLeuLeuHis 151
              :::||||||:::|||   |||||||||||||||||||||||||:::||||||||||||||
Db        246 CGTGGAAGAGAACCTCTGTTTGTTCCTTGTACACCAAAGGGATGCATAGAGCTACTGCAC 305

Qy        152 ArgTyrGlyValGluIleLysGlyLysArgAlaValValIleGlyArgSerAsnIleVal 171
              ||||||   |||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AGATACAATGTTTCTATTAAAGGAAAGAGGGCTGTTGTGATTGGTCGGAGCAATATTGTT 365

Qy        172 GlyMetProAlaAlaLeuLeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSer 191
              ||||||||||||||||||||||||||:::   :::||||||||||||||||||||||||||
Db        366 GGAATGCCAGCTGCTCTCTTGCTTCAAAGGGAAGATGCTACTGTCAGTATTGTCCATTCT 425

Qy        192 AsnThrLysLysProGluGluIleThrArgGlnAlaAspIleValIleAlaAlaValGly 211
              |||      ||||||||||||   ||||||||||||||:::||||||||||||||||
Db        426 AGAACCAGTAACCCCGAAGAGATCATAAGACAGGCAGATATTATCATTGCTGCTGTTGGG 485

Qy        212 ValAlaAsnLeuValArgGlySerTrpIleLysProGlyAlaAlaIleIleAspValGly 231
              ||||||::|||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        486 CAAGCAAACATGGTGAGGGGAAGCTGGATAAAGCCTGGTGCAGTCATTATTGATGTTGGA 545

Qy        232 IleAsnProValAspAspProGluSerProArgGlyTyrArgLeuValGlyAspValCys 251
              ||||||||||||:::||||||   ||||||||||||||:::||||||||||||||||||||
Db        546 ATCAACCCGGTAGAGGATCCAAATAGCCCCCGAGGTTACAAACTGGTGGGAGATGTTTGT 605

Qy        252 TyrGluGluAlaSerLysIleAlaGlyLeuIleThrProValProGlyGlyValGlyPro 271
              ||||||||||||   :::||||||      :::||||||||||||||||||||||||||||
Db        606 TATGAAGAAGCCATAAGAATTGCCTCTGCTGTTACACCAGTTCCTGGAGGAGTTGGTCCA 665

Qy        272 MetThrIleAlaMetLeuLeuSerAsnThrLeuGluSerAlaLysArgIleHisLysPhe 291
              ||||||||||||||||||||   |||||||||   ||||||||||||:::|||   |||
Db        666 ATGACCATAGCAATGCTTCTACAAAATACACTCATCTCTGCAAAGAGGGTGCACAATTTT 725

Qy        292 Lys 292
              :::
Db        726 GAA 728

RESULT 5
US-10-723-061-19
; Sequence 19, Application US/10723061
; Publication No. US20040132150A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV1
;  CURRENT APPLICATION NUMBER: US/10/723,061
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: US 09/903,814
;  PRIOR FILING DATE: 2001-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
;   LENGTH: 926
;   TYPE: DNA
;   ORGANISM: Glycine max
US-10-723-061-19

Alignment Scores:
Pred. No.:              3.35e-120       Length:         926
Score:                  1030.00         Matches:        194
Percent Similarity:     89.63%          Conservative:   22
Best Local Similarity:  80.50%          Mismatches:     25
Query Match:            69.31%          Indels:         0
DB:                     17              Gaps:           0

US-10-723-061-18 (1-292) x US-10-723-061-19 (1-926)

Qy         52 LysLysLysAlaCysGluAlaValGlyIleLysSerTyrGluValAsnLeuProGluAsp 71
              :::|||||||||||||||:::|||||||||   |||   |||   ||||||||||||:::
Db          6 GAGAAGAAGGCTTGTGAATCTGTTGGAATCAATTCTTTGGAAGCAAATCTGCCTGAGAAT 65

Qy         72 SerSerGluAspGluValLeuLysHisIleAlaThrPheAsnSerAspProSerValHis 91
              |||:::|||:::||||||||||   :::||||||   :::|||   ||||||||||||||
Db         66 TCCACAGAAGAAGAAGTTTTGAACTATATTGCAGGCTACAATGATGATCCTTCAGTTCAT 125

Qy         92 GlyIleLeuValGlnLeuProLeuProHisHisMetAsnAspGluAsnIleLeuAsnAla 111
              |||||||||||||||||||||||||   |||||||||::::::||||||||||||||||
Db        126 GGCATCCTCGTTCAGTTACCCTTACCTTCGCATATGAATGAGCAGAACATCTTGAACGCT 185

Qy        112 ValSerIleGluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMet 131
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 GTCAGGATTGAGAAGGATGTAGATGGTTTTCATCCGTTAAATATTGGTCGTCTTGCCATG 245

Qy        132 GlnGlyArgAspProPhePheValProCysThrProLysGlyCysMetGluLeuLeuHis 151
              :::||||||:::|||   |||||||||||||||||||||||||:::||||||||||||||
Db        246 CGTGGAAGAGAACCTCTGTTTGTTCCTTGTACACCAAAGGGATGCATAGAGCTACTGCAC 305

Qy        152 ArgTyrGlyValGluIleLysGlyLysArgAlaValValIleGlyArgSerAsnIleVal 171
              ||||||   |||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AGATACAATGTTTCTATTAAAGGAAAGAGGGCTGTTGTGATTGGTCGGAGCAATATTGTT 365

Qy        172 GlyMetProAlaAlaLeuLeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSer 191
              ||||||||||||||||||||||||||:::   :::||||||||||||||||||||||||||
Db        366 GGAATGCCAGCTGCTCTCTTGCTTCAAAGGGAAGATGCTACTGTCAGTATTGTCCATTCT 425

Qy        192 AsnThrLysLysProGluGluIleThrArgGlnAlaAspIleValIleAlaAlaValGly 211
              |||      ||||||||||||   ||||||||||||||:::||||||||||||||||
Db        426 AGAACCAGTAACCCCGAAGAGATCATAAGACAGGCAGATATTATCATTGCTGCTGTTGGG 485

Qy        212 ValAlaAsnLeuValArgGlySerTrpIleLysProGlyAlaAlaIleIleAspValGly 231
              ||||||:::|||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        486 CAAGCAAACATGGTGAGGGGAAGCTGGATAAAGCCTGGTGCAGTCATTATTGATGTTGGA 545

Qy        232 IleAsnProValAspAspProGluSerProArgGlyTyrArgLeuValGlyAspValCys 251
              ||||||||||||:::||||||   ||||||||||||||:::||||||||||||||||||||
Db        546 ATCAACCCGGTAGAGGATCCAAATAGCCCCCGAGGTTACAAACTGGTGGGAGATGTTTGT 605

Qy        252 TyrGluGluAlaSerLysIleAlaGlyLeuIleThrProValProGlyGlyValGlyPro 271
              ||||||||||||   :::||||||      :::||||||||||||||||||||||||||||
Db        606 TATGAAGAAGCCATAAGAATTGCCTCTGCTGTTACACCAGTTCCTGGAGGAGTTGGTCCA 665

Qy        272 MetThrIleAlaMetLeuLeuSerAsnThrLeuGluSerAlaLysArgIleHisLysPhe 291
              ||||||||||||||||||||   |||||||||   ||||||||||||:::|||   |||
Db        666 ATGACCATAGCAATGCTTCTACAAAATACACTCATCTCTGCAAAGAGGGTGCACAATTTT 725

Qy        292 Lys 292
              :::
```

```
Db        726 GAA 728

RESULT 6
US-10-425-114-33029
; Sequence 33029, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 33029
;   LENGTH: 1066
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMFLMO17055A08_FLI
US-10-425-114-33029

Alignment Scores:
Pred. No.:              1.79e-119      Length:          1066
Score:                  1025.00        Matches:         192
Percent Similarity:     81.53%         Conservative:    42
Best Local Similarity:  66.90%         Mismatches:      53
Query Match:            68.98%         Indels:          0
DB:                     13             Gaps:            0

US-10-723-061-18 (1-292) x US-10-425-114-33029 (1-1066)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||:::|||||||||||||||   :::|||     :::|||   |||:::|||   :::
Db         25 GCGCAGATCATCGATGGCAAGGCCATCGCCGCCGACGTCCGCCGCGAGGTCGCCGCCGAT 84

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::|||   :::     |||   |||:::||||||||||||||||||:::::||||||
Db         85 GTGGCCGCGCTCTCGTCGGCCCACGGACTCGTGCCGGGGCTGGCCGTGGTCATCGTGGGG 144

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||     |||:::||||||||     ||||||
Db        145 AGCAGGAAGGACTCGCAGACGTACGTGAACATGAAGCGCAAGGCGTGCGCCGAGGTCGGC 204

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||   |||   :::|||:::||||||||||||   ||||||      ::::::
Db        205 ATCTGCTCCATCGACGTCGACCTCCCGGAGGACATCTCCGAGACCGCGCTCGTCGCCGAG 264

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::         |||:::||||||::||||||||||||||||||||||||||||||||
Db        265 GTTCATCGCCTCAACGCTGACCCCGCCGTGCACGGGATCCTTGTCCAGCTTCCACTTCCT 324

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                 |||:::|||:::|||   ||||||:::   :::||||||||||||||||||||||||
Db        325 AAGCATATCAACGAAGAGAAGATACTGAGCGAGATTTCCATCGAGAAAGATGTGGATGGC 384

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||||||:::||||||||:::||||||:::|||   |||||||||
Db        385 TTCCATCCTCTCAACATTGGCAAGCTTGCAATGAAAGGCAGAGAGCCACTGTTCGTACCA 444

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||||||||||||   |||   ||||||   :::|||||||||
Db        445 TGTACGCCAAAGGGGTGCATGGAGCTCTTGTCAAGGAGCGGAGTCACTGTTAAAGGTAAG 504

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||:::||||||||||||||||||||:::|||   :::||||||||||
Db        505 CGGGCAGTTGTGGTTGGTCGCAGCAACATCGTCGGGCTACCTGTATCCCTGCTCCTTCTG 564

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::|||||||||||||:::|||||||||   |||       ||||||   |||
Db        565 AAGGCAGATGCGACCGTATCTGTTGTGCACTCGCGGACCCCTGATCCTGAAAGCATTGTA 624

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||||||   |||   |||   :::::::::|||   |||
Db        625 CGCGAAGCCGACATAGTCATCGCGGCAGCTGGGCAGGCTATGATGATCAAAGGTGACTGG 684

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||||:::||||||||||||   |||   ||||||||||||
Db        685 ATCAAGCCAGGTGCTGCGGTCATCGATGTCGGGACGAACTCCGTCGACGACCCTACCCGG 744

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                  ||||||||||||||||||||||||||   :::       |||||||||:::||||||
Db        745 AAGTCCGGGTACCGGCTCGTCGGCGATGTGGATTTCGCAGCGGCGAGCAAGGTTGCTGGG 804

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
                 :::|||||||||||||||||||||||||||||||:::||||||||||||   |||
Db        805 TACCTGACTCCGGTTCCCGGAGGCGTTGGCCCAATGACGGTGGCAATGCTGCTGAAGAAC 864

Qy        281 ThrLeuGluSerAlaLysArg 287
              |||::::::   |||||||||
Db        865 ACGGTGGATGGGGCAAAGCGG 885

RESULT 7
US-10-425-114-23607
; Sequence 23607, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 23607
;   LENGTH: 1415
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3597-067-D5_FLI
US-10-425-114-23607

Alignment Scores:
Pred. No.:              2.76e-119      Length:          1415
Score:                  1025.00        Matches:         192
Percent Similarity:     81.53%         Conservative:    42
Best Local Similarity:  66.90%         Mismatches:      53
Query Match:            68.98%         Indels:          0
DB:                     13             Gaps:            0

US-10-723-061-18 (1-292) x US-10-425-114-23607 (1-1415)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||:::|||||||||||||||   :::|||     :::|||   |||:::|||   :::
Db        406 GCGCAGATCATCGATGGCAAGGCCATCGCCGCCGACGTCCGCCGCGAGGTCGCCGCCGAT 465

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::|||   :::     |||   |||:::||||||||||||||||||:::::||||||
Db        466 GTGGCCGCGCTCTCGTCGGCCCACGGACTCGTGCCGGGGCTGGCCGTGGTCATCGTGGGG 525

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||     |||:::||||||||     ||||||
Db        526 AGCAGGAAGGACTCGCAGACGTACGTGAACATGAAGCGCAAGGCGTGCGCCGAGGTCGGC 585

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||   |||   :::|||:::||||||||||||   ||||||      ::::::
```

```
Db        586 ATCTGCTCCATCGACGTCGACCTCCCGGAGGACATCTCCGAGACCGCGCTCGTCGCCGAG 645

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100

Db        646 GTTCATCGCCTCAACGCTGACCCCGCCGTGCACGGGATCCTTGTCCAGCTTCCACTTCCT 705

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120

Db        706 AAGCATATCAACGAAGAGAAGATACTGAGCGAGATTTCCATCGAGAAAGATGTGGATGGC 765

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140

Db        766 TTCCATCCTCTCAACATTGGCAAGCTTGCAATGAAAGGCAGAGAGCCACTGTTCGTACCA 825

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160

Db        826 TGTACGCCAAAGGGGTGCATGGAGCTCTTGTCAAGGAGCGGAGTCACTGTTAAAGGTAAG 885

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180

Db        886 CGGGCAGTTGTGGTTGGTCGCAGCAACATCGTCGGGCTACCTGTATCCCTGCTCCTTCTG 945

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200

Db        946 AAGGCAGATGCGACCGTATCTGTTGTGCACTCGCGGACCCCTGATCCTGAAAGCATTGTA 1005

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220

Db       1006 CGCGAAGCCGACATAGTCATCGCGGCAGCTGGGCAGGCTATGATGATCAAAGGTGACTGG 1065

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240

Db       1066 ATCAAGCCAGGTGCTGCGGTCATCGATGTCGGGACGAACTCCGTCGACGACCCTACCCGG 1125

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260

Db       1126 AAGTCCGGGTACCGGCTCGTCGGCGATGTGGATTTCGCAGCGGCGAGCAAGGTTGCTGGG 1185

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280

Db       1186 TACCTGACTCCGGTTCCCGGAGGCGTTGGCCCAATGACGGTGGCAATGCTGCTGAAGAAC 1245

Qy        281 ThrLeuGluSerAlaLysArg 287

Db       1246 ACGGTGGATGGGGCAAAGCGG 1266

RESULT 8
US-10-425-114-11296
; Sequence 11296, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 11296
;   LENGTH: 1255
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 701050070_FLI
US-10-425-114-11296

Alignment Scores:
Pred. No.:              1.32e-118       Length:         1255
```

```
Score:                  1019.00         Matches:        189
Percent Similarity:     81.66%          Conservative:   47
Best Local Similarity:  65.40%          Mismatches:     53
Query Match:            68.57%          Indels:         0
DB:                     13              Gaps:           0

US-10-723-061-18 (1-292) x US-10-425-114-11296 (1-1255)

Qy          3 IleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIleAla 22

Db        312 GTGCTTGATGGAAAGTTGATATCCATGGAAATCAGATCAAAAATAGCTGCTAAGGTAAGA 371

Qy         23 LysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySerArg 42

Db        372 CAAATGAAGAAAGGCCTAGGAAAAGTTCCTGGATTAGCTGTAATCTTAGTGGGCCAAAGA 431

Qy         43 LysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIleLys 62

Db        432 AGGGATTCTCAAACTTACGTTCGCAACAAGATAATGGCCTGTGAAGAAGTTGGAATCAAG 491

Qy         63 SerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAla 82

Db        492 TCTTTGGTGACTGAATTACCCACTGATTGTGCAGTAACAGATGTTCAAAATGCCATCATG 551

Qy         83 ThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHis 102

Db        552 AGATTTAACAAAGATCCATCCATTCATGGTATTCTTGTGCAACTTCCTCTACCGCAACAT 611

Qy        103 MetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHis 122

Db        612 CTAGATGAGGAAAAAGTTTTGGATGCTGTATGTCTAGAAAAAGATGTGGATGGTTTCCAT 671

Qy        123 ProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThr 142

Db        672 CCCCTTAATATGGGGAATCTTGCCATAAGAGGAAGGGAGCCACTGTTTACTCCTTGTACT 731

Qy        143 ProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAla 162

Db        732 CCAAAGGGCTGTATTGAGTTATTGATCAGATCAGGAGTCGAGATTATGGGAAAAAAAGCC 791

Qy        163 ValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLysAla 182

Db        792 GTAGTGATTGGAACAAGTAATATTGTGGGTTTGCCAGCATCCTTGTTATTGCAGAGACAC 851

Qy        183 AsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGln 202

Db        852 CATGCAACAGTCACTGTCATACATGCCTTCACACAAAACCCTGAACAGATCACCTCAGAA 911

Qy        203 AlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLys 222

Db        912 GCTGATATTGTAGTTTCAGCTGCTGGAGTGCCTAATTTGGTCCGTGGGAACTGGATAAAA 971

Qy        223 ProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerProArg 242

Db        972 CCTGGTGCAGCTGTGATCGATGTTGGTACAACTCCGGTTGAGGATCCTGGCTGCGAGGAT 1031

Qy        243 GlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIle 262

Db       1032 GGTTATCGCCTCGCAGGTGACGTATGCTATGAAGAGGCCGTAAAAGTGGCATCCATTATT 1091

Qy        263 ThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeu 282

Db       1092 ACTCCTGTGCCTGGCGGGGTTGGGCCTATGACCGTTGCTATGCTACTATGTAATACATTA 1151

Qy        283 GluSerAlaLysArgIleHisLysPhe 291

Db       1152 GATTCTGCAAAGCGCATGCTCAATTTT 1178

RESULT 9
US-10-424-599-114288
; Sequence 114288, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 114288
;   LENGTH: 1798
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_74210C.1
US-10-424-599-114288

Alignment Scores:
Pred. No.:              2.3e-118      Length:          1798
Score:                  1019.00       Matches:         189
Percent Similarity:     81.66%        Conservative:    47
Best Local Similarity:  65.40%        Mismatches:      53
Query Match:            68.57%        Indels:          0
DB:                     13            Gaps:            0

US-10-723-061-18 (1-292) x US-10-424-599-114288 (1-1798)

Qy          3 IleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIleAla 22
              ::::::|||||||||||||:::::::   :::|||||||   :::|||||||   ::::::
Db        312 GTGCTTGATGGAAAGTTGATATCCATGGAAATCAGATCAAAAATAGCTGCTAAGGTAAGA 371

Qy         23 LysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySerArg 42
              :::|||||||      :::|||   ||||||||||||||||||||||||||||||   |||
Db        372 CAAATGAAGAAAGGCCTAGGAAAAGTTCCTGGATTAGCTGTAATCTTAGTGGGCCAAAGA 431

Qy         43 LysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIleLys 62
              :::|||||||||||||||||||||||||||      |||||||||   ||||||||||||
Db        432 AGGGATTCTCAAACTTACGTTCGCAACAAGATAATGGCCTGTGAAGAAGTTGGAATCAAG 491

Qy         63 SerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAla 82
              |||              ||||||   |||   :::      :::|||          |||
Db        492 TCTTTGGTGACTGAATTACCCACTGATTGTGCAGTAACAGATGTTCAAAATGCCATCATG 551

Qy         83 ThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHis 102
                 ||||||   |||||||||:::||||||||||||||||||||||||||||||   |||
Db        552 AGATTTAACAAAGATCCATCCATTCATGGTATTCTTGTGCAACTTCCTCTACCGCAACAT 611

Qy        103 MetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHis 122
              ::::::::::|||   :::|||:::||||||   :::|||||||||||||||||||||||
Db        612 CTAGATGAGGAAAAAGTTTTGGATGCTGTATGTCTAGAAAAAGATGTGGATGGTTTCCAT 671

Qy        123 ProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThr 142
              |||||||||:::|||   ||||||:::::||||||:::|||   |||   ||||||||||
Db        672 CCCCTTAATATGGGGAATCTTGCCATAAGAGGAAGGGAGCCACTGTTTACTCCTTGTACT 731

Qy        143 ProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAla 162
              ||||||||||||:::||||||||||   |||   ||||||||||||   ||||||:::|||
Db        732 CCAAAGGGCTGTATTGAGTTATTGATCAGATCAGGAGTCGAGATTATGGGAAAAAAAGCC 791

Qy        163 ValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLysAla 182
              ||||||||||||   ||||||||||||||:::||||||:::||||||||||||:::
Db        792 GTAGTGATTGGAACAAGTAATATTGTGGGTTTGCCAGCATCCTTGTTATTGCAGAGACAC 851

Qy        183 AsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGln 202
              :::||||||||:::::::::|||:::   |||:::   |||||||:::||||||   :::
Db        852 CATGCAACAGTCACTGTCATACATGCCTTCACACAAAACCCTGAACAGATCACCTCAGAA 911

Qy        203 AlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLys 222
              ||||||||||||::::::|||   ||||||   ||||||||||||||||:::||||||||
Db        912 GCTGATATTGTAGTTTCAGCTGCTGGAGTGCCTAATTTGGTCCGTGGGAACTGGATAAAA 971

Qy        223 ProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerProArg 242
              ||||||||||||:::|||||||||||      ||||||:::||||||
Db        972 CCTGGTGCAGCTGTGATCGATGTTGGTACAACTCCGGTTGAGGATCCTGGCTGCGAGGAT 1031

Qy        243 GlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIle 262
              ||||||||||||   ||||||||||||||||||||||||   |||:::|||   :::|||
Db       1032 GGTTATCGCCTCGCAGGTGACGTATGCTATGAAGAGGCCGTAAAAGTGGCATCCATTATT 1091

Qy        263 ThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeu 282
              ||||||||||||||||||||||||||||||:::||||||||||||   |||||||||
Db       1092 ACTCCTGTGCCTGGCGGGGTTGGGCCTATGACCGTTGCTATGCTACTATGTAATACATTA 1151

Qy        283 GluSerAlaLysArgIleHisLysPhe 291
              :::||||||||||||:::      |||
Db       1152 GATTCTGCAAAGCGCATGCTCAATTTT 1178

RESULT 10
US-09-903-814-15
; Sequence 15, Application US/09903814
; Patent No. US20020102689A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB-1179-A
;  CURRENT APPLICATION NUMBER: US/09/903,814
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
;   LENGTH: 1076
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-903-814-15

Alignment Scores:
Pred. No.:              2.51e-118     Length:          1076
Score:                  1016.00       Matches:         190
Percent Similarity:     81.18%        Conservative:    43
Best Local Similarity:  66.20%        Mismatches:      54
Query Match:            68.37%        Indels:          0
DB:                     9             Gaps:            0

US-10-723-061-18 (1-292) x US-09-903-814-15 (1-1076)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||::|||||||||||||||   :::|||      :::|||   |||:::|||   :::
Db         32 GCGCAGATCATCGATGGCAAGGCCATCGCCGCCGACGTCCGCCGCGAGGTCGCCGCCGAT 91

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::|||   :::       |||   |||:::|||||||||||||||||:::::||||||
Db         92 GTGGCCGCGCTCTCGTCGGCCCACGGACTCGTGCCGGGGCTGGCCGTGGTCATCGTGGGG 151

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||      |||:::|||||||||      ||||||
Db        152 AGCAGGAAGGACTCGCAGACGTACGTGAACATGAAGCGCAAGGCGTGCGCCGAGGTCGGC 211

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||   |||   :::|||:::||||||||||   ||||||      ::::::
Db        212 ATCTGCTCCATCGACGTCGACCTCCCGGAGGACATCTCCGAGACCGCGCTCGTCGCCGAG 271

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::      |||:::|||||||:::||||||||||||||||||||||||||||||||||
Db        272 GTTCATCGCCTCAACGCTGACCCCGCAGTGCACGGGATCCTTGTCCAGCTTCCACTTCCT 331

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||:::|||:::|||   |||||||:::   :::||||||||||||||||||||||||||
Db        332 AAGCATATCAACGAAGAGAAGATACTGAGCGAGATTTCCATCGAGAAAGATGTGGATGGC 391
```

```
Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||||:::|||||||||:::||||||:::|||   ||||||||
Db        392 TTCCATCCTCTCAACATTGGCAAGCTTGCAATGAAAGGCAGAGAGCCACTGTTCGTACCA 451

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||||||||||||||   |||   ||||||   :::|||||||||
Db        452 TGTACGCCAAAGGGGTGCATGGAGCTCTTGTCAAGGAGCGGAGTCACTGTTAAAGGTAAG 511

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||:::|||||||||||||||||||:::|||   :::||||||||||
Db        512 CGGGCAGTTGTGGTTGGTCGCAGCAACATCGTCGGGCTACCTGTATCCCTGCTCCTTCTG 571

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::|||||||||||:::|||||||||   |||      ||||||   |||
Db        572 AAGGCAGATGCGACCGTATCTGTTGTGCACTCGCGGACCCCTGATCCTGAAAGCATTGTA 631

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||||   |||   |||   :::::::::|||   |||
Db        632 CGCGAAGCTGACATAGTCATCGCGGCAGCTGGGCAGGCTATGATGATCAAAGGTGACTGG 691

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||||:::||||||||||   |||   :::||||||||||
Db        692 ATCAAGCCAGGTGCTGCGGTCATCGATGTCGGGACGAACTCCATCGATGACCCTACCCGG 751

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                  |||   ||||||||||||||||||   :::      |||||||||:::||||||
Db        752 AAGTCCGGGGTACGGCTCGTCGGCGATGTGGATTTCGCAGCGGCGAGCAAGGTTGCTGGG 811

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
                 :::||||||||||||||||||||||||||||||||||||:::||||||||||||   |||
Db        812 TACCTGACTCCGGTTCCCGGAGGCGTTGGCCCAATGACGGTGGCAATGCTGCTGAAGAAC 871

Qy        281 ThrLeuGluSerAlaLysArg 287
              |||::::::   |||||||||
Db        872 ACGGTGGATGGGGCAAAGCGG 892

RESULT 11
US-10-723-061-15
; Sequence 15, Application US/10723061
; Publication No. US20040132150A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV1
;  CURRENT APPLICATION NUMBER: US/10/723,061
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: US 09/903,814
;  PRIOR FILING DATE: 2001-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
;   LENGTH: 1076
;   TYPE: DNA
;   ORGANISM: Zea mays
US-10-723-061-15

Alignment Scores:
Pred. No.:              2.51e-118       Length:         1076
Score:                  1016.00         Matches:        190
Percent Similarity:     81.18%          Conservative:   43
Best Local Similarity:  66.20%          Mismatches:     54
Query Match:            68.37%          Indels:         0
DB:                     17              Gaps:           0

US-10-723-061-18 (1-292) x US-10-723-061-15 (1-1076)
```

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||:::|||||||||||||||   :::|||      :::|||   |||:::|||   :::
Db         32 GCGCAGATCATCGATGGCAAGGCCATCGCCGCCGACGTCCGCCGCGAGGTCGCCGCCGAT 91

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::|||   :::      |||   |||:::||||||||||||||||||:::::||||||
Db         92 GTGGCCGCGCTCTCGTCGGCCCACGGACTCGTGCCGGGGCTGGCCGTGGTCATCGTGGGG 151

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||      ||:::|||||||||      ||||||
Db        152 AGCAGGAAGGACTCGCAGACGTACGTGAACATGAAGCGCAAGGCGTGCGCCGAGGTCGGC 211

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||   |||   :::|||:::||||||||||   ||||||      ::::::
Db        212 ATCTGCTCCATCGACGTCGACCTCCCGGAGGACATCTCCGAGACCGCGCTCGTCGCCGAG 271

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::      |||:::|||||:::||||||||||||||||||||||||||||||||||||
Db        272 GTTCATCGCCTCAACGCTGACCCCGCAGTGCACGGGATCCTTGTCCAGCTTCCACTTCCT 331

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                 |||:::|||:::|||   ||||||:::   :::||||||||||||||||||||||
Db        332 AAGCATATCAACGAAGAGAAGATACTGAGCGAGATTTCCATCGAGAAAGATGTGGATGGC 391

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||:::|||||||||:::|||||||:::|||   |||||||||
Db        392 TTCCATCCTCTCAACATTGGCAAGCTTGCAATGAAAGGCAGAGAGCCACTGTTCGTACCA 451

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||||||||||||||   |||   ||||||   :::|||||||||
Db        452 TGTACGCCAAAGGGGTGCATGGAGCTCTTGTCAAGGAGCGGAGTCACTGTTAAAGGTAAG 511

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||:::|||||||||||||||||||:::|||   :::||||||||||
Db        512 CGGGCAGTTGTGGTTGGTCGCAGCAACATCGTCGGGCTACCTGTATCCCTGCTCCTTCTG 571

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::|||||||||||:::|||||||||   |||      ||||||   |||
Db        572 AAGGCAGATGCGACCGTATCTGTTGTGCACTCGCGGACCCCTGATCCTGAAAGCATTGTA 631

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||||   |||   |||   :::::::::|||   |||
Db        632 CGCGAAGCTGACATAGTCATCGCGGCAGCTGGGCAGGCTATGATGATCAAAGGTGACTGG 691

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||||:::||||||||||   |||      :::||||||||||
Db        692 ATCAAGCCAGGTGCTGCGGTCATCGATGTCGGGACGAACTCCATCGATGACCCTACCCGG 751

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                  |||   ||||||||||||||||||   :::      |||||||||:::||||||
Db        752 AAGTCCGGGGTACGGCTCGTCGGCGATGTGGATTTCGCAGCGGCGAGCAAGGTTGCTGGG 811

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
                 :::||||||||||||||||||||||||||||||||||||:::||||||||||||   |||
Db        812 TACCTGACTCCGGTTCCCGGAGGCGTTGGCCCAATGACGGTGGCAATGCTGCTGAAGAAC 871

Qy        281 ThrLeuGluSerAlaLysArg 287
              |||::::::   |||||||||
Db        872 ACGGTGGATGGGGCAAAGCGG 892

RESULT 12
US-10-437-963-74482
; Sequence 74482, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
```

```
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 74482
;   LENGTH: 1356
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_74664C.1
US-10-437-963-74482

Alignment Scores:
Pred. No.:              1.15e-117       Length:         1356
Score:                  1012.00         Matches:        191
Percent Similarity:     80.84%          Conservative:   41
Best Local Similarity:  66.55%          Mismatches:     55
Query Match:            68.10%          Indels:         0
DB:                     17              Gaps:           0

US-10-723-061-18 (1-292) x US-10-437-963-74482 (1-1356)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||:::|||||||||||||||||   ||||||      ||||||   |||:::|||   :::
Db         99 GCGCAGATCATCGACGGGAAGGCGGTCGCCGCCGACATCCGCCGCGAGGTCGCCGCCGAC 158

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              .:::|||   :::      |||      :::||||||||||||||||||:::::||||||
Db        159 GTCGCCGCCCTCTCCTCCGCCCACAACCTCGTGCCCGGGCTGGCGGTGGTGATCGTGGGG 218

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||:::   |||:::|||||||||      ||||||
Db        219 AGCAGGAAGGACTCGCAGACGTACGTGCAGATGAAGCGCAAGGCGTGCGCCGAGGTCGGC 278

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||:::|||   :::|||:::|||   ||||||   ||||||      ::::::
Db        279 ATCCGCTCCGTCGACGTCGACCTCGCCGAGGACATCTCCGAGGCGGCGCTCGTCGCCGAG 338

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::         |||:::|||||||:::||||||||||||||||||||||||||||||||
Db        339 GTCCACCGCCTCAACGCCGACCCAGCCGTCCATGGCATACTTGTCCAGCTTCCATTGCCA 398

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||:::|||:::|||   |||||||||   :::|||:::|||||||||||||||||||||
Db        399 AAGCATATTAATGAGGAAAAGATATTGAACGAGATCAGCTTAGAGAAAGATGTTGATGGG 458

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||||:::||||||||||:::||||||||||.|   |||:::|||
Db        459 TTCCATCCTCTGAATATTGGCAAGCTTGCAATGAAAGGCAGAGACCCACTGTTCCTACCA 518

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||||||||||||||||||||||||||||   |||   ||||||   |||   ||||||
Db        519 TGCACGCCGAAGGGATGCATGGAGCTCCTAACACGGAGCGGAGTTACCATCAATGGGAAG 578

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||:::|||||||||||||||||||||:::|||   :::||||||||||
Db        579 CGAGCGGTCGTGGTTGGCCGCAGCAACATTGTCGGGCTACCTGTATCCCTGCTTCTTCTG 638

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::|||||||||||||||||||||||||   |||      ||||||   |||
Db        639 AAGGCGGATGCGACCGTATCCATTGTTCACTCCCGGACCCCAAACCCTGAAAGTATTGTC 698

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||   |||   |||   ::::::::::|||   |||
Db        699 CGTGAAGCAGACATTGTCATTGCAGCGGCTGGCCAGGCTATGATGATCAAAGGAGACTGG 758

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||||:::||||||||||||   |||   :::   ||||||
Db        759 ATCAAACCAGGCGCTGCTGTCATCGACGTTGGGACGAACTCCATCAGTGACCCAACCAGG 818

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||||||   :::   |||   ||||||:::||||||
Db        819 AAATCGGGGTACAGACTCGTCGGCGATGTGGATTTCGCAGAGGTGAGCAAGGTTGCTGGT 878

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              :::||||||||||||||||||||||||||||||||:::||||||||||||   |||
Db        879 CACCTGACTCCAGTCCCAGGTGGCGTTGGCCCCATGACGGTGGCGATGTTGCTGAAGAAC 938

Qy        281 ThrLeuGluSerAlaLysArg 287
              |||::::::   |||||||||
Db        939 ACGGTGGATGGAGCGAAACGT 959

RESULT 13
US-10-424-599-125693
; Sequence 125693, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 125693
;   LENGTH: 1286
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_8450C.1
US-10-424-599-125693

Alignment Scores:
Pred. No.:              1.9e-117        Length:         1286
Score:                  1010.00         Matches:        188
Percent Similarity:     81.88%          Conservative:   47
Best Local Similarity:  65.51%          Mismatches:     52
Query Match:            67.97%          Indels:         0
DB:                     13              Gaps:           0

US-10-723-061-18 (1-292) x US-10-424-599-125693 (1-1286)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::||||||||||||||   ||||||:::   ||||||   |||||||||   |||
Db         64 GCCACCGTAATCGACGGCAAAGCCGTTGCGCAAACCATCCGATCTGAAATCGCCGACGAG 123

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::   ::::::              |||   |||||||||||||||||:::::||||||
Db        124 GTGCGCCAGCTCTCTCAAAAATACGGCAAGGTTCCGGGATTAGCGGTGGTGATAGTAGGG 183

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              :::|||||||||||||||:::||||||      |||:::|||||||||      :::||||
Db        184 AACAGAAAAGATTCACAGAGCTATGTTGGAATGAAGAGAAAGGCATGCGCTGAATTGGGA 243

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::::::|||:::|||||||||      |||:::   |||::::::|||
Db        244 ATCAAATCCTTCGATGTGGACCTTCCCGAACAAGTCTCCCAAGCTGAACTAATCAAACAA 303

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::         |||:::::|||   ||||||||||||||||||||||||||||||||||
Db        304 GTTCACCAGTTGAATGCTAACCCTGACGTACATGGTATATTGGTTCAGCTCCCGTTGCCT 363

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||:::|||:::|||   :::|||      :::|||:::|||||||||||||||||||||
Db        364 AAGCACATAAATGAAGAGGAAGTTCTGACTGAAATCAGCCTTGAGAAGGATGTAGATGGT 423

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||||:::|||||||||:::|||||||||||   |||:::|||
Db        424 TTTCATCCTTTGAACATTGGCAAGCTTGCAATGAAAGGCAGAGACCCCCTGTTCCTTCCC 483
```

```
Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db        484 TGCACTCCCAAGGCATGTATTGAACTATTACAACGAAGTGGTGTAAGTATTAAGGGAAAG 543

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
Db        544 AAAGCCGTTGTGGTTGGTAGAAGCAACATAGTTGGATTACCAGCTTCATTGCTGCTTTTG 603

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db        604 AAAGCAGATGCCACAGTTACTATTATTCATTCACACACAAGTCAGCCAGAAAGTATCATA 663

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db        664 CGTGAAGCAGATATTGTTATTGCAGCAGCAGGACAGCCGATGATGATCAAGGGCAGTTGG 723

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db        724 ATCAAACCTGGAGCTGCAGTGATAGATGTTGGCACAAATGCTGTTGATGACCCAACAAAG 783

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db        784 AAGTCAGGTTACAGGCTTGTTGGAGATGTAGATTTTGGGGAAGCATCTAAAGTTGCTGGT 843

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db        844 TGGATTACTCCAGTTCCTGGTGGTGTAGGTCCAATGACAGTCACAATGTTGCTGAAGAAT 903

Qy        281 ThrLeuGluSerAlaLysArg 287
Db        904 ACTTTGGATGGAGCCAAGCGC 924

RESULT 14
US-10-425-114-9450
; Sequence 9450, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 9450
;   LENGTH: 1409
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700847483_FLI
US-10-425-114-9450

Alignment Scores:
Pred. No.:              2.32e-115      Length:         1409
Score:                  994.00         Matches:        187
Percent Similarity:     80.84%         Conservative:   45
Best Local Similarity:  65.16%         Mismatches:     55
Query Match:            66.89%         Indels:         0
DB:                     13             Gaps:           0

US-10-723-061-18 (1-292) x US-10-425-114-9450 (1-1409)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db         87 GCCACCGTAATCGACGGCAAAGCCGTTGCGCAAATCATCCGATCTGAAATCGCCGACGAG 146

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40

Db        147 GTGCGCCTGCTCTCTCAAAAATACGGCAAGGTTCCGGGACTAGCAGTTGTGATAGTAGGG 206

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db        207 AACAGAAAGGATTCACAAAGCTACGTGGGAATGAAGAGAAAGGCATGCGCCGAATTGGGA 266

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db        267 ATCAAATCCTTCGATGTGGACCTTCCCGAACAAGTCTCCCAAGCTGAACTAATAAAACAA 326

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db        327 GTTCACGAGTTGAATGTTAACCCTGATGTACATGGTATATTGGTTCAGCTTCCGTTGCCT 386

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db        387 AAGCACATAAATGAAGAGAAAGTTCTGACTGAAATCAGCCTTGAGAAGGATGTAGATGGC 446

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
Db        447 TTTCATCCTTTGAACATTGGCAAGCTTGCAATGAAAGGCAGAGACCCCCTATTCCTTCCC 506

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db        507 TGCACTCCCAAGGCATGTATTGAACTATTACAGCGAAGTGGTGTAAGTATAAAGGGAAAG 566

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
Db        567 AAAGCAGTTGTGGTTGGCAGAAGCAACATAGTTGGACTACCAGCTTCATTGCTGCTTTTG 626

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db        627 AAAGCAGATGCTACAGTTACTATTGTTCATTCACACACAAGTCAACCGGAAAGTATCATA 686

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db        687 CATGAAGCAGATATTGTTATTGCAGCAGCAGGACAGCCGATGATGATCAAGGGAAATTGG 746

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db        747 ATCAAACCAGGAGCTGCAGTCATAGATGTTGGCACAAATGCTGTTGATGACCCAACAAAG 806

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db        807 AAGTCGGGTTACAGGCTTGTTGGAGATGTAGATTTTGAGGAAGCATCTAAAGTTGCTGGT 866

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db        867 TGGATTACTCCTGTTCCTGGTGGTGTAGGTTCAATGACAGTGACAATGTTGCTGAAGAAT 926

Qy        281 ThrLeuGluSerAlaLysArg 287
Db        927 ACTGTGGAGGGAGCCAAGCGC 947

RESULT 15
US-10-437-963-23018
; Sequence 23018, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
```

; SEQ ID NO 23018
; LENGTH: 1413
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_28138C.1
US-10-437-963-23018

Alignment Scores:

Pred. No.:	1.34e-114	Length:	1413
Score:	988.00	Matches:	186
Percent Similarity:	80.77%	Conservative:	45
Best Local Similarity:	65.03%	Mismatches:	55
Query Match:	66.49%	Indels:	0
DB:	17	Gaps:	0

US-10-723-061-18 (1-292) x US-10-437-963-23018 (1-1413)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   |||||||||||||||    ||||||:::   ||||||   :::|||||||   |||
Db        324 GCGACCATTATTGATGGCAAGTCTGTCGCGGAAGACATAAGGTTTCAGATCGCTGAAGAA 383

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::   :::||||||:::|||:::|||   ||||||||||||||||||:::|||||||||
Db        384 GTTCGTCAAATGAAAAATGCAGTGGGGCATGTGCCTGGCCTAGCTGTTGTATTGGTCGGC 443

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
                 |||:::||||||::::::|||||||||   |||   |||   ||||||   ||||||
Db        444 GATAGAAGGGACTCTGAGTCATACGTGCGATACAAAATTAAGGGCTGTGAGGAGGTTGGA 503

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||            ||||||   :::   :::||||||   |||:::
Db        504 ATAAAGTCCTTGTTGGCGGAGTTGCCTGGGAACTGTACGGAAGACGTAGTGGTGGATTCT 563

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::   ||||||   |||||||||||||||||||||||||||||||||||||||||
Db        564 GTGTCAAGATTCAATGAAGATCCATCTGTTCATGGCATTCTTGTGCAACTACCTCTACCA 623

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                |||||||::::::|||   ||||||:::|||:::|||:::|||||||||||||||||
Db        624 CAGCATATGGATGAAGAAAGGATTCTGAGCGCTATTAGCCTAGAGAAGGATGTTGATGGT 683

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||:::|||   ||||||::::::   |||   |||   |||||||||
Db        684 TTCCATCCATTGAATGTTGGTAATCTTGCCCTCAGAAGCCGGAAGCCTTTATTTGTGCCC 743

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||       |||   |||:::|||||||||   :::   |||:::|||:::   ||||||
Db        744 TGTGCTGCAAAGGCTTGCCTTGAGTTATTGCTCCAGTCTGGGATTGAACTCATGGGGAAA 803

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
                  ||||||||||||||   :::||||||:::|||   :::||||||||||||||
Db        804 CATGTCACTGTGATTGGGAGAAGCAAAGTTGTTGGGTTGCCCACTTCTTTACTTTTACAG 863

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::|||||||||||||||:::|||:::   |||       ||||||||||||||
Db        864 AGACACCATGCCACTGTTAGTATTATCCATGCTTTCACAACAAATCCAGAAGAGATTACC 923

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              ||||||:::|||||||||||||:::|||   ||||||||||||||||||||||||||||||
Db        924 CGTCAATCTGATATTGTGATCTCAGCTGCTGGAGTGGCCAATCTTGTAAGAGGAAGCTGG 983

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              :::|||   ||||||   :::|||||||||||   ||||||::::::||||||   |||
Db        984 TTGAAGAAAGGTGCAGTTGTGATCGATGTTGGAACAAACCCAATTGAGGATCCAACCAGT 1043

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                    ||||||||||||   ||||||||||||:::|||||||||   |||:::|||
Db       1044 GATTACGGCTATCGATTAACTGGAGATGTCTGCTTCGAAGAAGCTGTGAAGTTGGCCTCT 1103

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
                 |||||||||||||||||||||||||||||:::|||||||||||||||||||:::|||
Db       1104 GCTATAACTCCAGTTCCTGGTGGTGTGGGACCGGTGACCATTGCGATGCTTCTTGCGAAT 1163
```

```
Qy        281 ThrLeuGluSerAlaLys 286
                 |||||||:::|||||||||
Db       1164 ACACTTGACTCAGCGAAA 1181
```

Search completed: July 24, 2004, 11:12:47
Job time : 587 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 24, 2004, 06:18:40 ; Search time 59 Seconds
(without alignments)
476.067 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1238	83.3	360	2	T01226	probable methylene
2	1034	69.6	294	2	T50664	methylenetetrahydr
3	1017	68.4	350	2	G84807	methylenetetrahydr
4	1009	67.9	310	2	T01224	probable methylene
5	763.5	51.4	284	2	AD1607	methylenetetrahydr
6	755.5	50.8	284	2	AH1244	methylenetetrahydr
7	749.5	50.4	937	2	T40723	c-1-tetrahydrofola
8	741	49.9	319	2	AE3573	methenyltetrahydro
9	739.5	49.8	946	1	A29550	methylenetetrahydr
10	735	49.5	935	1	A31903	methylenetetrahydr
11	730	49.1	344	2	C87400	FolD bifunctional
12	724	48.7	285	2	D95095	hypothetical prote
13	718	48.3	935	1	A35367	methylenetetrahydr
14	716	48.2	279	2	H83997	methylenetetrahydr
15	715	48.1	306	2	A97963	methenyltetrahydro
16	713.5	48.0	975	1	A28174	methylenetetrahydr
17	712	47.9	933	1	S53523	methylenetetrahydr
18	707.5	47.6	294	2	E86731	hypothetical prote
19	704	47.4	284	2	F82031	methylenetetrahydr
20	689	46.4	282	2	G81358	methylenetetrahydr
21	683.5	46.0	283	2	E69626	methylenetetrahydr
22	681.5	45.9	282	1	A64081	methylenetetrahydr
23	679	45.7	288	1	JS0662	methylenetetrahydr
24	677	45.6	284	2	H83421	5,10-methylene-tet
25	676	45.5	285	2	S71924	methylenetetrahydr
26	674	45.4	288	2	G90702	5,10-methylene-tet
27	674	45.4	288	2	B85553	5,10-methylene-tet
28	673	45.3	288	1	S36633	methylenetetrahydr
29	673	45.3	288	2	AF0569	hypothetical prote
30	668	45.0	291	2	F70463	methylenetetrahydr
31	668	45.0	299	2	AD1833	methylenetetrahydr
32	667	44.9	299	2	AH2648	hypothetical prote
33	667	44.9	319	2	G97430	hypothetical prote
34	665.5	44.8	286	2	E89875	FolD bifunctional
35	662	44.5	969	2	T40147	probable tetrahydr
36	661	44.5	311	2	H82136	methylenetetrahydr
37	660	44.4	288	2	AB0344	methenyltetrahydro
38	652.5	43.9	357	1	S32562	methylenetetrahydr
39	648	43.6	297	2	A84296	methylenetetrahydr
40	632	42.5	315	2	A71288	probable methylene
41	629.5	42.4	350	1	A33267	methylenetetrahydr
42	627.5	42.2	306	2	T51705	methylenetetrahydr
43	626	42.1	295	2	C82558	bifunctional methy
44	623.5	42.0	344	1	DEHUMT	methylenetetrahydr
45	621	41.8	294	2	S76287	5,10-methylene-tet

ALIGNMENTS

RESULT 1
T01226
probable methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) F6N23.26 - Arabidops
N;Alternate names: protein F6N23.26
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 03-Jun-2002
C;Accession: T01226
R;Geisel, C.
submitted to the EMBL Data Library, April 1998
A;Description: The sequence of A. thaliana F6N23.
A;Reference number: Z14281
A;Accession: T01226
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-360 <GEI>
A;Cross-references: EMBL:AF058919; NID:g3047100; PID:g3047116; GSPDB:GN00063; ATSP:F6N23.
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: ATSP:F6N23.26
A;Map position: 5
A;Introns: 61/1; 169/3; 248/3; 304/3
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate
C;Keywords: NADP; oxidoreductase
F;71-354/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>

Query Match 83.3%; Score 1238; DB 2; Length 360;
Best Local Similarity 79.1%; Pred. No. 1.1e-90;
Matches 231; Conservative 36; Mismatches 25; Indels 0; Gaps 0;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :|||| |||:||:|| :|:::||::|||:|||||||| |||| ||||||||||::||
Db         69 AIVIDGKAVAKKIRDEITIEVSRMKESIGVIPGLAVILVGDRKDSATYVRNKKKACDSVG 128

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||:|| | |||||:|||| :: || |||||||||||||| ||:::|||||||||||||
Db        129 IKSFEVRLAEDSSEEEVLKSVSGFNDDPSVHGILVQLPLPSHMDEQNILNAVSIEKDVDG 188

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||||||:||:| |||||||||:|||||| :|||||||||||||||||||||||||
Db        189 FHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNIEIKGKRAVVIGRSNIVGMPAALLLQ 248

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              : :|||||:|| || ||||||:|||:|:|||  |:|||||||||| :|||||||:|| :
Db        249 REDATVSIIHSRTKNPEEITREADIIISAVGQPNMVRGSWIKPGAVLIDVGINPVEDPSA 308

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
              ||||||||:|||||||:|  |||||||||||||||||||| |||||| |:
Db        309 ARGYRLVGDICYEEASKVASAITPVPGGVGPMTIAMLLSNTLTSAKRIHNFQ 360
```

RESULT 2
T50664

methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) / methenyltetrahydrofolate c
C;Species: Pisum sativum (garden pea)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 03-Jun-2002
C;Accession: T50664
R;Chen, L.; Nargang, N.E.; Cossins, E.A.
submitted to the EMBL Data Library, October 1997
A;Description: A cDNA sequence encoding pea leaf cytosolic bifunctional 5,10-methylenete
A;Reference number: Z25166
A;Accession: T50664
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-294 <CHE>
A;Cross-references: EMBL:AF030516; PIDN:AAD01907.1
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate
C;Keywords: hydrolase; multifunctional enzyme; NADP; oxidoreductase

  Query Match 69.6%; Score 1034; DB 2; Length 294;
  Best Local Similarity 67.2%; Pred. No. 1.3e-74;
  Matches 193; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

Qy 1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
Db 2 ATVIDGKAVAQTIRSEIADEVRLLSQKYGKVPGLAVVIVGNRKDSQSYVGMKRKACAELG 61

Qy 61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
Db 62 IKSFDIDLPEDASEAEIIKNVHELNANPDVHGILVQLPLPKHVNEEKVLTEISISKDVDG 121

Qy 121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
Db 122 FHPLNIGKLAMKGRDPLFLPCTPKACLELLSRSGVSIKGKKAVVVGRSNIVGLPASLLL 181

Qy 181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
Db 182 KADATVTIVHSHTSQPETIIREADIVIAAAGQAKMIKGSWIKPGAAVIDVGTNSVDDPTR 241

Qy 241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
Db 242 KSGYRLVGDVDFEEASKVAGWITPVPGGVGPMTVTMLLKNTLEGAKR 288

RESULT 3
G84807
methylenetetrahydrofolate dehydrogenase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
C;Accession: G84807
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84807
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-350 <STO>
A;Cross-references: GB:AE002093; NID:g3786006; PIDN:AAC67352.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g38660
A;Map position: 2
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

  Query Match 68.4%; Score 1017; DB 2; Length 350;
  Best Local Similarity 67.5%; Pred. No. 3.7e-73;
  Matches 193; Conservative 36; Mismatches 57; Indels 0; Gaps 0;

Qy 3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
Db 63 VIDGNVIAEEIRTKIISEVGKMKKAVGKVPGLAVVLVGEQRDSQTYVRNKIKACEETGIK 122

Qy 63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
Db 123 SVLAELPEDCTEGQIISVLRKFNEDTSIHGILVQLPLPQHLNESKILNMVRLEKDVDGFH 182

Qy 123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
Db 183 PLNVGNLAMRGREPLFVSCTPKGCVELLIRTGVEIAGKNAVVIGRSNIVGLPMSLLLQRH 242

Qy 183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
Db 243 DATVSTVHAFTKDPEHITRKADIVIAAAGIPNLVRGSWLKPGAVVIDVGTTPVEDSSCEF 302

Qy 243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
Db 303 GYRLVGDVCYEEALGVASAITPVPGGVGPMTITMLLCNTLEAAKRI 348

RESULT 4
T01224
probable methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) F6N23.28 - Arabidops
N;Alternate names: protein F6N23.28
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 03-Jun-2002
C;Accession: T01224
R;Geisel, C.
submitted to the EMBL Data Library, April 1998
A;Description: The sequence of A. thaliana F6N23.
A;Reference number: Z14281
A;Accession: T01224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-310 <GEI>
A;Cross-references: EMBL:AF058919; NID:g3047100; PID:g3047121; GSPDB:GN00063; ATSP:F6N23.
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: ATSP:F6N23.28
A;Map position: 5
A;Introns: 87/1; 119/3; 198/3; 254/3
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate
C;Keywords: NADP; oxidoreductase
F;23-304/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>

  Query Match 67.9%; Score 1009; DB 2; Length 310;
  Best Local Similarity 66.4%; Pred. No. 1.3e-72;
  Matches 194; Conservative 33; Mismatches 29; Indels 36; Gaps 1;

Qy 1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
Db 55 AIVIDGKAEAKKIRDDIKIEVSRMKESIGVVPA---------- 87

Qy 61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
Db 88 ---------EDSSEEEVLKYVSGFNDDPSVHGVLVQLPLPSHMDEQNILNAVSIEKDVDG 138

Qy 121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
Db 139 FHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNIEFKGKRAVVIGRSNIVGMPAALLLQ 198

Qy 181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
Db 199 KEDATVSIIHSRTMNPEELTRQADILISAVGKPNMVRGSWIKPGAVLIDVGIKPVEDPSA 258

Qy 241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
Db 259 AGGERLVGDICYVEASKIASAITPVPGDVGPMTIAMLLSNTLTSAKRIHNFQ 310

RESULT 5
AD1607
methylenetetrahydrofolate dehydrogenase and methenyltetrahydrofolate cyclohydrolase homol
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AD1607
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;

D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AD1607
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-284 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC96628.1; PID:g16413870; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: folD
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

Query Match 51.4%; Score 763.5; DB 2; Length 284;
Best Local Similarity 52.2%; Pred. No. 3.7e-53;
Matches 151; Conservative 59; Mismatches 70; Indels 9; Gaps 4;

```
Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              :|||||  :||:|:|::  |:|::    |  |||||:|||  : |:||||||:|  |  |:
Db          3 EIIDGKKLAKEIQEKVTSEVAELVKQ-GKKPGLAVVLVGDNQASRTYVRNKQKRTEEAGM 61

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
              ||  : |||  :|:::|| :   | | ::||||||||||| |:::| :::|:| :||||||
Db         62 KSVLIELPETVTEEKLLKVVEELNEDNTIHGILVQLPLPKHISEEKVIDAISFDKDVDGF 121

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              ||:|:|  | : |:|  ||||||  | :||:    | :|:||||||||||||||| | | ||
Db        122 HPVNVGNLFI-GKDS-FVPCTPAGIIELIKSTGTQIEGKRAVVIGRSNIVGKPVAQLLLN 179

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              ||||:|  || ||   :: ::|||:: | |:|  |:  :|||||  :||||:      :
Db        180 ENATVTIAHSRTKDLPQVAKEADILVVATGLAKFVKKEYIKPGAIVIDVGM------DRD 233

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHK 290
                :| ||| :::  : || ||||||||||||| |||:|||::|||| |
Db        234 ENNKLCGDVDFDDVKEQAGFITPVPGGVGPMTITMLLANTLKAAKRIWK 282
```

RESULT 6
AH1244
methylenetetrahydrofolate dehydrogenase and methenyltetrahydrofolate cyclohydrolase homo
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AH1244
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1244
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-284 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99438.1; PID:g16410776; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: folD
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

Query Match 50.8%; Score 755.5; DB 2; Length 284;
Best Local Similarity 51.6%; Pred. No. 1.6e-52;
Matches 149; Conservative 60; Mismatches 71; Indels 9; Gaps 4;

```
Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              :|||||  :||:|:|::  |:|::    |  |||||:|||  : |:||||||:|  |  |:
Db          3 EIIDGKKLAKEIQEKVTREVAELVKE-GKQPGLAVVLVGDNQASRTYVRNKQKRTEEAGM 61

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
              ||  : |||: :|:::|  :   | | ::||||||||||| |:::| ::: :| :||||||
Db         62 KSVLIELPENVTEEKLLSVVEELNEDKTIHGILVQLPLPEHISEEKVIDTISFDKDVDGF 121

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              ||:|:|  | : |:|  ||||||  | :||:    | :|:||||||||||||||| | | ||
Db        122 HPVNVGNLFI-GKDS-FVPCTPAGIIELIKSTGTQIEGKRAVVIGRSNIVGKPVAQLLLN 179

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              ||||:|  || ||   :: ::|||:: | |:|  |:  :|||||  :||||:      :
Db        180 ENATVTIAHSRTKDLPQVAKEADILVVATGLAKFVKKDYIKPGAIVIDVGM------DRD 233

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHK 290
                :| ||| :::  : || ||||||||||||| |||:|||::|||| |
Db        234 ENNKLCGDVDFDDVVQEAGFITPVPGGVGPMTITMLLANTLKAAKRIWK 282
```

RESULT 7
T40723
c-1-tetrahydrofolate synthase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C;Accession: T40723
R;Lyne, M.; Purnelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Saunde
submitted to the EMBL Data Library, July 1999
A;Reference number: Z21946
A;Accession: T40723
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-937 <LYN>
A;Cross-references: EMBL:AL096796; PIDN:CAB46709.1; GSPDB:GN00067; SPDB:SPBC839.16
A;Experimental source: strain 972h-; cosmid c839
C;Genetics:
A;Gene: SPDB:SPBC839.16
A;Map position: 2
A;Introns: 12/2; 60/3
C;Superfamily: C1-tetrahydrofolate synthase; formate-tetrahydrofolate ligase homology; m
F;312-937/Domain: formate-tetrahydrofolate ligase homology <FTL>

Query Match 50.4%; Score 749.5; DB 2; Length 937;
Best Local Similarity 50.0%; Pred. No. 2.3e-51;
Matches 147; Conservative 50; Mismatches 90; Indels 7; Gaps 2;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVP----GLAVILVGSRKDSQTYVRNKKKAC 56
              | :::|  :|:::|||:  :|: :|    | |      | :| || |:|| ||| | :|
Db          2 ALLLEGTSLARKVREELREQISSIK---SVDPYFNVSLKIIQVGGREDSNVYVRMKTRAA 58

Qy         57 EAVGIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEK 116
                ||    || ||| :| ::|  |  || ||:||||:|||||| |:|:: |  ||: ||
Db         59 NEAGISCEHVNFPEDITEYDLLLAIKGFNEDPTVHGIIVQLPLPAHINEQIITEAVAPEK 118

Qy        117 DVDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAA 176
              |||||   |:|:|  :   | |  |||| | :|  ||: ::|| |||||||||| | :
Db        119 DVDGFCETNLGKLTKREGQPLFTACTPKGIMCILKHYGINVQGKHAVVIGRSNIVGRPMS 178

Qy        177 LLLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVD 236
              :||:||||||:: || |:   :| | ||||:||:|: : |:  |:| |   ||||||  :
Db        179 ILLEKANATVTLCHSKTESIADIVRTADIVVAAIGIPHFVKADWLKKGVVAIDVGINSIP 238

Qy        237 DPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHK 290
              |      |||| ||: :| | ::|  ||||| |||||:|||| | :||| |  |
Db        239 DATKKSGYRLTGDIDFENAKEVASAITPVPGSVGPMTVAMLLQNVVESAVRFRK 292
```

RESULT 8
AE3573
methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9) [imported] - Brucella melitensis (st
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C;Accession: AE3573
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, N
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesso
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AE3573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-319 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL53752.1; PID:g17984679; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEII0510
A;Map position: II
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

Query Match 49.9%; Score 741; DB 2; Length 319;
Best Local Similarity 51.6%; Pred. No. 2.6e-51;
Matches 149; Conservative 49; Mismatches 89; Indels 2; Gaps 2;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |::|||| :|: :   :  |  |:  | |||||:||::||    || || :| :  :  |
Db         22 AQLIDGKKLAEDVVSTVKTETEKLVAATGVVPGIAVVIVGEDPASQVYVASKSRKAKECG 81

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
                | : :||| :|| |:|  |   |:||::|||||||||| |::   ::  :: ||||||
Db         82 FHSVQHDLPETASEQELLNLIEGLNNDPAIHGILVQLPLPGHIDSGRVIQTIAPEKDVDG 141

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHR-YGVEIKGKRAVVIGRSNIVGMPAALLL 179
              || :|:|:|     :  |||||| | | :: | :| :: |  ||||||||||| |   ||
Db        142 FHFINVGKLGTGEVETAFVPCTPAGAMIMIERVHGRDLSGLNAVVIGRSNIVGKPMFNLL 201

Qy        180 QKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPE 239
               |||||:: || ||    | | |||::||||   :|:| |:|||| :|||||| :  ||
Db        202 LAANATVTVAHSRTKDLPAIARNADILVAAVGRPQMVKGDWVKPGATVIDVGINRIPAPE 261

Qy        240 SPRG-YRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                |  |||||| : || |:|| |||||||||||||||::||| :| |
Db        262 RGEGKTRLVGDVDFAEAEKVAGAITPVPGGVGPMTIAMLMANTLTAACR 310
```

RESULT 9
A29550
methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) - yeast (Saccharomyces cerev
N;Alternate names: C1-tetrahydrofolate synthase; protein G7733; protein YGR204w
N;Contains: formate-tetrahydrofolate ligase (EC 6.3.4.3); methenyltetrahydrofolate cyclo
C;Species: Saccharomyces cerevisiae
C;Date: 21-May-1988 #sequence_revision 05-Apr-1995 #text_change 19-Jul-2002
C;Accession: A29550; S53927; S64526; S63853
R;Staben, C.; Rabinowitz, J.C.
J. Biol. Chem. 261, 4629-4637, 1986
A;Title: Nucleotide sequence of the Saccharomyces cerevisiae ADE3 gene encoding C1-tetra
A;Reference number: A92580; MUID:86168166; PMID:3514599
A;Accession: A29550
A;Molecule type: DNA
A;Residues: 1-946 <STA>
A;Cross-references: EMBL:M12878; NID:g171004; PIDN:AAA66316.1; PID:g171005
R;Guerreiro, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the EMBL Data Library, April 1995
A;Description: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII
es, of the human.
A;Reference number: S53922
A;Accession: S53927
A;Molecule type: DNA
A;Residues: 1-946 <GUE>
A;Cross-references: EMBL:Z49133; NID:g790489; PIDN:CAA88997.1; PID:g790495
A;Experimental source: strain S228C
R;Guerreiro, P.; Barreiros, T.; Cyrne, L.; Soares, H.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64517
A;Accession: S64526
A;Molecule type: DNA
A;Residues: 1-946 <GUW>
A;Cross-references: EMBL:Z72989; NID:g1323364; PIDN:CAA97231.1; PID:g1323365; GSPDB:GN00
A;Experimental source: strain S288C

R;Guerreiro, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Maia e Silva, A.; Rodrigues-Pousac
Yeast 12, 273-280, 1996
A;Title: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII reveals
terial electron-transferring flavoproteins (beta-chain) and of the Escherichia coli phosp
A;Reference number: S63848; MUID:97060019; PMID:8904340
A;Accession: S63853
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-946 <GUF>
A;Cross-references: EMBL:Z49133; NID:g790489; PIDN:CAA88997.1; PID:g790495
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C;Genetics:
A;Gene: SGD:ADE3; MIPS:YGR204w
A;Cross-references: SGD:S0003436; MIPS:YGR204w
A;Map position: 7R
C;Superfamily: C1-tetrahydrofolate synthase; formate-tetrahydrofolate ligase homology; me
C;Keywords: hydrolase; ligase; multifunctional enzyme; NADP; one-carbon metabolism; oxidc
F;5-297/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>
F;322-946/Domain: formate-tetrahydrofolate ligase homology <FTL>

Query Match 49.8%; Score 739.5; DB 1; Length 946;
Best Local Similarity 50.5%; Pred. No. 1.4e-50;
Matches 149; Conservative 45; Mismatches 92; Indels 9; Gaps 2;

```
Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAI-GVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              :::|||  |:| |  || ||  ::  : |  | ||:| ||:| || |||| |:|| |  |
Db          4 QVLDGKACAQQFRSNIANEIKSIQGHVPGFAPNLAIIQVGNRPDSATYVRMKRKAAEEAG 63

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
               | :  ::| | ::| |||:::   | ||  |||:|||||| |:::: | : |  ||||||
Db         64 IVANFIHLDESATEFEVLRYVDQLNEDPHTHGIIVQLPLPAHLDEDRITSRVLAEKDVDG 123

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              | | ||| |  :   |||:|||||| :|||:  | |:| |:|||||:||| | | ||:
Db        124 FGPTNIGELNKKNGHPFFLPCTPKGIIELLHKANVTIEGSRSVVIGRSDIVGSPVAELLK 183

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKP--------GAAIIDVGI 232
               |:||:| || |:        ||||: |:|    |:| | ||           :||||
Db        184 SLNSTVTITHSKTRDIASYLHDADIVVVAIGQPEFVKGEWFKPRDGTSSDKKTVVIDVGT 243

Qy        233 NPVDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
              | | ||    |:: |||| : || |   |||||||||||:|||: ||| :|||
Db        244 NYVADPSKKSGFKCVGDVEFNEAIKYVHLITPVPGGVGPMTVAMLMQNTLIAAKR 298
```

RESULT 10
A31903
methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) - human
N;Contains: formate-tetrahydrofolate ligase (EC 6.3.4.3); methylenetetrahydrofolate dehyc
C;Species: Homo sapiens (man)
C;Date: 21-May-1990 #sequence_revision 05-Apr-1995 #text_change 19-Jul-2002
C;Accession: A31903
R;Hum, D.W.; Bell, A.W.; Rozen, R.; MacKenzie, R.E.
J. Biol. Chem. 263, 15946-15950, 1988
A;Title: Primary structure of a human trifunctional enzyme. Isolation of a cDNA encoding
ase.
A;Reference number: A31903; MUID:89034046; PMID:3053686
A;Accession: A31903
A;Molecule type: mRNA
A;Residues: 1-935 <HUM>
A;Cross-references: GB:J04031; NID:g187464; PIDN:AAA59574.1; PID:g307178
C;Genetics:
A;Gene: GDB:MTHFD1; MTHFD; MTHFC
A;Cross-references: GDB:120704; OMIM:172460
A;Map position: 14q24-14q24
C;Superfamily: C1-tetrahydrofolate synthase; formate-tetrahydrofolate ligase homology; me
C;Keywords: hydrolase; ligase; multifunctional enzyme; NADP; one-carbon metabolism; oxidc
F;6-292/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>
F;318-935/Domain: formate-tetrahydrofolate ligase homology <FTL>

Query Match 49.5%; Score 735; DB 1; Length 935;
Best Local Similarity 49.0%; Pred. No. 3.2e-50;
Matches 145; Conservative 56; Mismatches 91; Indels 4; Gaps 3;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAI-GVVPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              |:|::|| :: |||  :  :: ::|: : |  | ||:: ||:| ||  |:  | || | :
Db          4 AEILNGKEISAQIRARLKNQVTQLKEQVPGFTPRLAILQVGNRDDSNLYINVKLKAAEEI 63

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPL--PHHMNDENILNAVSIEKD 117
              |||:  : ||  ::| ||:|:| : | | :||| ||||||   : :| | ::||:: |||
Db         64 GIKATHIKLPRTTTESEVMKYITSLNEDSTVHGFLVQLPLDSENSINTEEVINAIAPEKD 123

Qy        118 VDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAAL 177
              |||   :| ||||    :  |:|||||||:||:   || | |: |||:||| ||| | |
Db        124 VDGLTSINAGRLARGDLNDCFIPCTPKGCLELIKETGVPIAGRHAVVVGRSKIVGAPMHD 183

Qy        178 LLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDD 237
              ||   ||||:  || |   :|   : ||:: | |   :|:| |||||| :|| ||| | |
Db        184 LLLWNNATVTTCHSKTAHLDEEVNKGDILVVATGQPEMVKGEWIKPGAIVIDCGINYVPD 243

Qy        238 PESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR-IHKFK 292
               : | | ::|||| |:|| : |  ||||||||||||:|||: :|:|||| : |||
Db        244 DKKPNGRKVVGDVAYDEAKERASFITPVPGGVGPMTVAMLMQSTVESAKRFLEKFK 299
```

RESULT 11
C87400
FolD bifunctional protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 27-Nov-2001
C;Accession: C87400
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-344 <STO>
A;Cross-references: GB:AE005673; NID:g13422543; PIDN:AAK23199.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1217
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

  Query Match 49.1%; Score 730; DB 2; Length 344;
  Best Local Similarity 51.7%; Pred. No. 2.2e-50;
  Matches 151; Conservative 47; Mismatches 88; Indels 6; Gaps 3;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              ||:|||| | :| :|| |:| :|   || ||||||:|||     || ||||| :   | |
Db         51 AKLIDGKAFAADLRAKIAEEVAALKAEHGVTPGLAVVLVGEDPASQVYVRNKGEQTTAAG 110

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              : |    ||| :::||:|  :|  |:|| :||:||| |:|  |::  ::: |:| :|||||
Db        111 MYSETHRLPETTTQDELLAVVAKLNADPKIHGVLVQFPVPPHISQMDVVAALSPDKDVDG 170

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
               | ||||     |  |||| ||| |:     ::||| ||||||::| | | :|
Db        171 LTVTNAGRLA--SGLPALAPCTPTGCMMLIRDAIGDLKGKTAVVIGRSNLMGKPMAQMLL 228

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGIN--PVDDP 238
               |: ||:| || ::    | ||||||:|||| | :|:  |:|||| :|||||   |  ||
Db        229 AADCTVTIAHSRSQDLPSIVRQADIVVAAVGRAEMVKADWVKPGAVVIDVGITRFPARDP 288

Qy        239 ESPRG--YRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |:      |||||| ::|  ::|| |||||||||||||| ||:||| :|||:
Db        289 EAAAAGKTRLVGDVAFDEVREVAGAITPVPGGVGPMTIACLLANTLTAAKRL 340
```

RESULT 12
D95095
hypothetical protein SP0825 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C;Accession: D95095
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, E nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: D95095
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-285 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK74957.1; PID:g14972298; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0825
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

  Query Match 48.7%; Score 724; DB 2; Length 285;
  Best Local Similarity 50.0%; Pred. No. 5.1e-50;
  Matches 143; Conservative 54; Mismatches 81; Indels 8; Gaps 2;

```
Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              :|||| :| ::: ::| : ||:|:  |:|||| ||||     || |||||:::  | |
Db          3 QIIDGKALAAKLQGQLAEKTAKLKEETGLVPGLVVILVGDNPASQVYVRNKERSALAAGF 62

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
              :|  | :||  :::|:|  || :| ||: ||||||||| |:::| :|  |:  |||||||
Db         63 RSEVVRVPETITQEELLDLIAKYNQDPAWHGILVQLPLPKHIDEEAVLLAIDPEKDVDGF 122

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              ||||:|||       |  :| || | ||: | ||:::::|| |||||||||| | | ||
Db        123 HPLNMGRL--WSGHPVMIPSTPAGIMEMFHEYGIDLEGKNAVVIGRSNIVGKPMAQLLLA 180

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              ||||:: || |    ::  :|||:: |:| |  |   ::|||| :||||:|  ::
Db        181 KNATVTLTHSRTHNLSKVAAKADILVVAIGRAKFVTADFVKPGAVVIDVGMNRDENG--- 237

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                :| ||| ||  : :|  |||||||||||||| ||:  | ::| |
Db        238 ---KLCGDVDYEAVAPLASHITPVPGGVGPMTITMLMEQTYQAALR 280
```

RESULT 13
A35367
methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) - rat
N;Contains: formate-tetrahydrofolate ligase (EC 6.3.4.3); methylenetetrahydrofolate dehyc
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Aug-1990 #sequence_revision 05-Apr-1995 #text_change 19-Jul-2002
C;Accession: A35367
R;Thigpen, A.E.; West, M.G.; Appling, D.R.
J. Biol. Chem. 265, 7907-7913, 1990
A;Title: Rat C-1-tetrahydrofolate synthase. cDNA isolation, tissue-specific levels of the
A;Reference number: A35367; MUID:90243660; PMID:2186031
A;Accession: A35367
A;Molecule type: mRNA
A;Residues: 1-935 <THI>
A;Cross-references: GB:J05519
C;Superfamily: C1-tetrahydrofolate synthase; formate-tetrahydrofolate ligase homology; me
C;Keywords: hydrolase; ligase; multifunctional enzyme; NADP; one-carbon metabolism; oxidc
F;6-292/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>
F;318-935/Domain: formate-tetrahydrofolate ligase homology <FTL>

  Query Match 48.3%; Score 718; DB 1; Length 935;
  Best Local Similarity 48.0%; Pred. No. 7.1e-49;
  Matches 142; Conservative 56; Mismatches 94; Indels 4; Gaps 3;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAI-GVVPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              | |::||:|: |||  :  :  :|:: : |  ||||:: || | ||  |:  | || : :
Db          4 AGILNGKVVSAQIRNLLKTQFTQMQEQVPGFTPGLAILQVGDRDDSNLYINVKLKAAQEI 63

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPL--PHHMNDENILNAVSIEKD 117
              |||:  : ||  |:| ||||:: : | | :||| :|||||   : :| | ::||:: |||
```

```
Db          64 GIKATHIKLPRTSTESEVLKYVISLNEDATVHGFIVQLPLDSENSINTEAVINAIAPEKD 123

Qy         118 VDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAAL 177
               |||   :| |:||       |:|||||||:||:   ||:| |: |||:||| ||| |
Db         124 VDGLTSINAGKLARGDLKDCFIPCTPKGCLELIKETGVQIAGRHAVVVGRSKIVGAPMHD 183

Qy         178 LLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDD 237
               ||   ||||:  || |   ::   : ||:: | |   :|:| |||||| :|| ||| | |
Db         184 LLLWNNATVTTCHSKTADLDKEVNKGDILVVATGQPEMVKGEWIKPGAVVIDCGINYVPD 243

Qy         238 PESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR-IHKFK 292
                 | ::|||| |:|| : |  ||||||||||||:|||: :|:|||:  : |||
Db         244 DTKPNGRKVVGDVAYDEAKEKASFITPVPGGVGPMTVAMLMQSTVESAQLFLKKFK 299


RESULT 14
H83997
methylenetetrahydrofolate dehydrogenase folD [imported] - Bacillus halodurans (strain C-
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: H83997
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: H83997
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-279 <STO>
A;Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BAB06503.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: folD
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

  Query Match             48.2%;  Score 716;  DB 2;  Length 279;
  Best Local Similarity   50.7%;  Pred. No. 2.1e-49;
  Matches  145;  Conservative   52;  Mismatches   79;  Indels   10;  Gaps    4;

Qy           1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
               | ::||| :| : |:|:  |: |:|:  |: |||||||||:   |  |||:|:|||| :|
Db           3 AVLLDGKALAAKKRDELKSEVTKLKER-GITPGLAVILVGNDPASTVYVRSKQKACEQIG 61

Qy          61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
               | |    ||  :||:|:|  |   |:||::|||||||||  :::: ::  :|  |||||
Db          62 IYSVLKELPASTSEEELLTEIDRLNNDPTIHGILVQLPLPEQISEQAVIERISPAKDVDG 121

Qy         121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
               |||:::||: | | |  |:|||| | : :|    ||| ||  .||:|||||| |   |:
Db         122 FHPISVGRM-MIGEDT-FLPCTPFGVLVMLQEANVEIAGKHVVVVGRSNIVGKPVGQLML 179

Qy         181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
                 :|||:  || ||   |||:||||:| ||| |  :  | :| || :||||||  ||
Db         180 NEHATVTYCHSRTKNLTEITKQADILIVAVGRARFIDASHVKEGAVVIDVGINRVDG--- 236

Qy         241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAK 286
                 :| ||| :|   ::|  :|||||||||||| |||:||:::||
Db         237 ----KLCGDVDFESVREVASHLTPVPGGVGPMTITMLLANTIQAAK 278


RESULT 15
A97963
methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9) folD [imported] - Streptococcus pne
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C;Accession: A97963
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: A97963
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-306 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK99533.1; PID:g15458321; GSPDB:GN00174
C;Genetics:
A;Gene: folD
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate
C;Keywords: hydrolase

  Query Match             48.1%;  Score 715;  DB 2;  Length 306;
  Best Local Similarity   49.7%;  Pred. No. 2.9e-49;
  Matches  142;  Conservative   54;  Mismatches   82;  Indels    8;  Gaps    2;

Qy           2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
               :||||| :| ::: ::| : ||:|:   :|||| |||||    || |||||:::  | |
Db          24 QIIDGKALAAKLQGQLAEKTAKLKEETALVPGLVVILVGDNPASQVYVRNKERSALAAGS 83

Qy          62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
               :|  | :||  :::|:|  || :| ||: |||||||||| |:::| :| |:  |||||||
Db          84 RSEVVRVPETITQEELLDLIAKYNQDPAWHGILVQLPLPKHIDEEAVLLAIDPEKDVDGF 143

Qy         122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
               ||||:|||       |  :| || | ||: | ||:::|| ||||||||||| | | ||
Db         144 HPLNMGRL--WSGHPVMIPSTPAGIMEMFHEYGIDLEGKNAVVIGRSNIVGKPMAQLLLA 201

Qy         182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
                ||||:: || |    ::  :|||:: |:| |  |   ::|||| :||||:|  ::
Db         202 KNATVTLTHSRTHNLAKVAAKADILVVAIGRAKFVTADFVKPGAVVIDVGMNRDENG--- 258

Qy         242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                  :| ||| ||  : :|  |||||||||||||| ||:  | ::| |
Db         259 ---KLCGDVDYEAVAPLASHITPVPGGVGPMTITMLMEQTYQAALR 301


Search completed: July 24, 2004, 06:40:09
Job time : 60 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 24, 2004, 04:13:15 ; Search time 52 Seconds
(without alignments)
292.394 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	762	51.3	934	1	C1TC_DROME	O96553 d c-1-tetra
2	739.5	49.8	946	1	C1TC_YEAST	P07245 s c-1-tetra
3	738	49.7	934	1	C1TC_MOUSE	Q922d8 m c-1-tetra
4	735	49.5	934	1	C1TC_HUMAN	P11586 h c-1-tetra
5	731	49.2	934	1	C1TC_RAT	P27653 r c-1-tetra
6	713.5	48.0	975	1	C1TM_YEAST	P09440 s c-1-tetra
7	692	46.6	284	1	FOLD_STRTR	P96050 streptococc
8	683.5	46.0	283	1	FOLD_BACSU	P54382 bacillus su
9	681.5	45.9	281	1	FOLD_HAEIN	P44313 haemophilus
10	679	45.7	287	1	FOLD_ECOLI	P24186 escherichia
11	676	45.5	285	1	FOLD_PHOPO	P51696 photobacter
12	673	45.3	287	1	FOLD_SALTI	Q60006 salmonella
13	669	45.0	287	1	FOLD_SALTY	P58688 salmonella
14	652.5	43.9	357	1	MTDC_DROME	Q04448 d bifunctio
15	629.5	42.4	350	1	MTDC_MOUSE	P18155 m bifunctio
16	623.5	42.0	344	1	MTDC_HUMAN	P13995 h bifunctio
17	608	40.9	286	1	FOLD_CHLPN	Q9z8k3 chlamydia p
18	602	40.5	290	1	FOLD_HELPY	P56467 helicobacte
19	597	40.2	290	1	FOLD_HELPJ	Q9zlq4 helicobacte
20	589	39.6	285	1	FOLD_BUCAI	P57557 buchnera ap
21	585.5	39.4	287	1	FOLD_CHLMU	Q9pkw1 chlamydia m
22	576.5	38.8	287	1	FOLD_CHLTR	O84081 chlamydia t
23	573.5	38.6	310	1	FOLD_RICPR	Q9zd32 rickettsia
24	564	38.0	285	1	FOLD_BUCAP	Q8k979 buchnera ap
25	418	28.1	276	1	FOLD_THEAC	Q05213 thermoplasm
26	300	20.2	273	1	FOLD_MYCGE	P47259 mycoplasma
27	286	19.2	269	1	FOLD_MYCPN	P75096 mycoplasma
28	230	15.5	320	1	MTD1_YEAST	Q02046 saccharomyc
29	106	7.1	342	1	OCD2_RHIME	P58339 rhizobium m
30	104	7.0	372	1	DHA2_STAAM	Q931p7 staphylococ
31	103	6.9	372	1	DHA2_STAAN	Q99tf4 staphylococ
32	103	6.9	372	1	DHA2_STAAW	Q8nw54 staphylococ
33	98	6.6	266	1	AROE_HALN1	Q9hs68 halobacteri
34	98	6.6	371	1	DHA_STAEP	Q8cnw8 staphylococ
35	97.5	6.6	230	1	SUMT_METIV	P29564 methanobact
36	96	6.5	463	1	YE11_METJA	Q58806 methanococc
37	95	6.4	283	1	AROE_METTH	O26344 methanobact
38	94	6.3	297	1	SVFA_BACSU	Q04809 bacillus su
39	94	6.3	378	1	DHA_BACSU	Q08352 bacillus su
40	92.5	6.2	373	1	MDHM_CHLRE	Q42686 chlamydomon
41	92	6.2	502	1	COAT_BPT6	Q38055 bacteriopha
42	92	6.2	1025	1	YEGO_ECOLI	P76399 escherichia
43	91.5	6.2	320	1	OCD_RHIME	P33728 rhizobium m
44	91	6.1	447	1	MCRB_METJA	Q58252 methanococc
45	90.5	6.1	372	1	DHA_BACST	P17557 bacillus st

ALIGNMENTS

RESULT 1
C1TC_DROME
ID C1TC_DROME STANDARD; PRT; 934 AA.
AC O96553; Q8T0P2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE C-1-tetrahydrofolate synthase, cytoplasmic (C1-THF synthase)
DE [Includes: Methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5);
DE Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9);
DE Formyltetrahydrofolate synthetase (EC 6.3.4.3)].
GN PUG OR CG4067.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=99054704; PubMed=9832531;
RA Rong Y.S., Golic K.G.;
RT "Dominant defects in Drosophila eye pigmentation resulting from a
RT euchromatin-heterochromatin fusion gene.";
RL Genetics 150:1551-1566(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

```
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE OF 569-933 FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Head;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +
CC       phosphate + 10-formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Present in all tissues.
CC   -!- MISCELLANEOUS: This trifunctional enzyme consists of two major
CC       domains: a N-terminal part, containing the methylene-THF
CC       dehydrogenase and the methenyl-THF cyclohydrolase activities and a
CC       larger formyl-THF synthetase domain.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; TO OTHER DEHYDROGENASE/
CC       CYCLOHYDROLASE ENZYMES OR DOMAINS.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG TO OTHER
CC       FORMYLTETRAHYDROFOLATE SYNTHETASE DOMAINS OR ENZYMES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF082097; AAC78847.1; -.
DR   EMBL; AE003688; AAG22140.1; -.
DR   EMBL; AY069146; AAL39291.1; ALT_INIT.
DR   HSSP; P11586; 1A4I.
DR   FlyBase; FBgn0020385; pug.
DR   InterPro; IPR000559; Fmtethyd_synth.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF01268; FTHFS; 1.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00721; FTHFS_1; 1.
DR   PROSITE; PS00722; FTHFS_2; FALSE_NEG.
DR   PROSITE; PS00766; THF_DHG_CYH_1; FALSE_NEG.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Hydrolase; Ligase; Oxidoreductase; NADP; Multifunctional enzyme;
KW   One-carbon metabolism; ATP-binding; Purine biosynthesis;
KW   Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis.
FT   DOMAIN        1    304       METHYLENETETRAHYDROFOLATE DEHYDROGENASE
FT                                AND CYCLOHYDROLASE (BY SIMILARITY).
FT   DOMAIN      305    934       FORMYLTETRAHYDROFOLATE SYNTHETASE (BY
FT                                SIMILARITY).
FT   NP_BIND     379    386       ATP (BY SIMILARITY).
SQ   SEQUENCE   934 AA;  99626 MW;  42CC88A69A4D5345 CRC64;
```

```
  Query Match           51.3%;  Score 762;  DB 1;  Length 934;
  Best Local Similarity  55.2%;  Pred. No. 6.3e-51;
  Matches  160;  Conservative   38;  Mismatches   88;  Indels    4;  Gaps    3;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKM-KDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              |||| |  ||| ||||:  |:  | |      |||| :: || |:||  |:| | ||   :
Db          4 AKIISGTAVAKSIREELRNEVTAMSKQLADFVPGLRIVQVGGREDSNVYIRMKIKAATEI 63

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPL--PHHMNDENILNAVSIEKD 117
              || :  | ||   :| |:|  |   | || ||||:||:||     ::    | :||| |||
Db         64 GIDAAHVQLPRSITEVELLDKINDLNEDPRVHGIIVQMPLDCDTPIDSHRITDAVSPEKD 123

Qy        118 VDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAAL 177
              ||| | :| ||||: |    |:|||| ||:||: | |||| | ||||:||| ||| |||
Db        124 VDGLHTVNEGRLAI-GDLGGFLPCTPWGCLELIRRSGVEIAGARAVVLGRSKIVGTPAAE 182

Qy        178 LLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDD 237
              ||: |||||:: || |:  |||| |||::  :||| :|:|||||||| :|| |||   |
Db        183 LLKWANATVTVCHSKTRNLEEITRSADILVVGIGVAEMVKGSWIKPGAVVIDCGINVKPD 242

Qy        238 PESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                | :|||| | || ::|| :||||||||||:|||: ||: || |
Db        243 ASKASGSKLVGDVDYAEALQVAGHLTPVPGGVGPMTVAMLMKNTVRSAAR 292
```

```
RESULT 2
C1TC_YEAST
ID   C1TC_YEAST     STANDARD;      PRT;   946 AA.
AC   P07245;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   C-1-tetrahydrofolate synthase, cytoplasmic (C1-THF synthase)
DE   [Includes: Methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5);
DE   Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9);
DE   Formyltetrahydrofolate synthetase (EC 6.3.4.3)].
GN   ADE3 OR YGR204W OR G7733.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86168166; PubMed=3514599;
RA   Staben C., Rabinowitz J.C.;
RT   "Nucleotide sequence of the Saccharomyces cerevisiae ADE3 gene
RT   encoding C1-tetrahydrofolate synthase.";
RL   J. Biol. Chem. 261:4629-4637(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=97060019; PubMed=8904340;
RA   Guerreiro P., Barreiros T., Soares H., Cyrne L., Maia e Silva A.,
RA   Rodrigues-Pousada C.;
RT   "Sequencing of a 17.6 kb segment on the right arm of yeast chromosome
RT   VII reveals 12 ORFs, including CCT, ADE3 and TR-I genes, homologues of
RT   the yeast PMT and EF1G genes, of the human and bacterial electron-
RT   transferring flavoproteins (beta-chain) and of the Escherichia coli
RT   phosphoserine phosphohydrolase, and five new ORFs.";
RL   Yeast 12:273-280(1996).
RN   [3]
RP   MUTAGENESIS.
RX   MEDLINE=89247401; PubMed=2541774;
RA   Barlowe C.K., Williams M.E., Rabinowitz J.C., Appling D.R.;
RT   "Site-directed mutagenesis of yeast C1-tetrahydrofolate synthase:
RT   analysis of an overlapping active site in a multifunctional enzyme.";
RL   Biochemistry 28:2099-2106(1989).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +
```

```
CC       phosphate + 10-formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- MISCELLANEOUS: This trifunctional enzyme consists of two major
CC       domains: a N-terminal part, containing the methylene-THF
CC       dehydrogenase and the methenyl-THF cyclohydrolase activities and a
CC       larger formyl-THF synthetase domain.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; TO OTHER DEHYDROGENASE/
CC       CYCLOHYDROLASE ENZYMES OR DOMAINS.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG TO OTHER
CC       FORMYLTETRAHYDROFOLATE SYNTHETASE DOMAINS OR ENZYMES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M12878; AAA66316.1; -.
DR   EMBL; Z49133; CAA88997.1; -.
DR   EMBL; Z72989; CAA97231.1; -.
DR   PIR; A29550; A29550.
DR   HSSP; P11586; 1A4I.
DR   GermOnline; 141516; -.
DR   SGD; S0003436; ADE3.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0009113; P:purine base biosynthesis; IMP.
DR   InterPro; IPR000559; Fmtethyd_synth.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF01268; FTHFS; 1.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00721; FTHFS_1; 1.
DR   PROSITE; PS00722; FTHFS_2; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Hydrolase; Ligase; Oxidoreductase; NADP; Multifunctional enzyme;
KW   One-carbon metabolism; ATP-binding; Purine biosynthesis;
KW   Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis.
FT   DOMAIN        1    319       METHYLENETETRAHYDROFOLATE DEHYDROGENASE
FT                                AND CYCLOHYDROLASE.
FT   DOMAIN      320    946       FORMYLTETRAHYDROFOLATE SYNTHETASE.
FT   NP_BIND     384    391       ATP (BY SIMILARITY).
FT   MUTAGEN      11     11       C->S: ALMOST NO EFFECT ON ACTIVITY.
FT   MUTAGEN     144    144       C->S: CYCLOHYDROLASE ACTIVITY IS REDUCED
FT                                20-FOLD; DEHYDROGENASE KM IS INCREASED 7-
FT                                FOLD.
FT   MUTAGEN     257    257       C->S: CYCLOHYDROLASE ACTIVITY IS
FT                                INCREASED 2-FOLD. DEHYDROGENASE KM IS
FT                                INCREASED 2-FOLD.
SQ   SEQUENCE   946 AA;  102205 MW;  BC2897EC380CBA6F CRC64;

  Query Match             49.8%;  Score 739.5;  DB 1;  Length 946;
  Best Local Similarity   50.5%;  Pred. No. 3.4e-49;
  Matches  149;  Conservative   45;  Mismatches   92;  Indels    9;  Gaps    2;

Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAI-GVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              :::|||  |:| | || ||  ::  : |  | ||:| ||:| || |||| |:|| |  |
Db          4 QVLDGKACAQQFRSNIANEIKSIQGHVPGFAPNLAIIQVGNRPDSATYVRMKRKAAEEAG 63

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | :  ::| | ::| |||:::  | ||  |||:|||||| |:::: | : |  ||||||
Db         64 IVANFIHLDESATEFEVLRYVDQLNEDPHTHGIIVQLPLPAHLDEDRITSRVLAEKDVDG 123

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              | | ||| |  :   |||:|||||| :||||:  | |:| |:|||||:||| | | ||:

Db        124 FGPTNIGELNKKNGHPFFLPCTPKGIIELLHKANVTIEGSRSVVIGRSDIVGSPVAELLK 183

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKP--------GAAIIDVGI 232
                |:||:| || |:        ||||: |:|    |:| | ||           :||||
Db        184 SLNSTVTITHSKTRDIASYLHDADIVVVAIGQPEFVKGEWFKPRDGTSSDKKTVVIDVGT 243

Qy        233 NPVDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
              | | ||    |:: |||| : || |   |||||||||||||:|||: ||| :|||
Db        244 NYVADPSKKSGFKCVGDVEFNEAIKYVHLITPVPGGVGPMTVAMLMQNTLIAAKR 298


RESULT 3
C1TC_MOUSE
ID   C1TC_MOUSE     STANDARD;      PRT;   934 AA.
AC   Q922D8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   C-1-tetrahydrofolate synthase, cytoplasmic (C1-THF synthase)
DE   [Includes: Methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5);
DE   Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9);
DE   Formyltetrahydrofolate synthetase (EC 6.3.4.3)].
GN   MTHFD1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +
CC       phosphate + 10-formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- MISCELLANEOUS: This trifunctional enzyme consists of two major
CC       domains: a N-terminal part, containing the methylene-THF
CC       dehydrogenase and the methenyl-THF cyclohydrolase activities and a
CC       larger formyl-THF synthetase domain.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; TO OTHER DEHYDROGENASE/
CC       CYCLOHYDROLASE ENZYMES OR DOMAINS.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG TO OTHER
CC       FORMYLTETRAHYDROFOLATE SYNTHETASE DOMAINS OR ENZYMES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; BC008523; AAH08523.1; -.
DR   MGD; MGI:1342005; Mthfd1.
DR   InterPro; IPR000559; Fmtethyd_synth.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF01268; FTHFS; 1.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00721; FTHFS_1; 1.
DR   PROSITE; PS00722; FTHFS_2; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Hydrolase; Ligase; Oxidoreductase; NADP; Multifunctional enzyme;
KW   One-carbon metabolism; ATP-binding; Purine biosynthesis;
KW   Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    304       METHYLENETETRAHYDROFOLATE DEHYDROGENASE
FT                                AND CYCLOHYDROLASE.
FT   DOMAIN      305    934       FORMYLTETRAHYDROFOLATE SYNTHETASE.
FT   NP_BIND     379    386       ATP (BY SIMILARITY).
SQ   SEQUENCE   934 AA;  101124 MW;  DC5C4FE95FC59F19 CRC64;

  Query Match             49.7%;  Score 738;  DB 1;  Length 934;
  Best Local Similarity   49.0%;  Pred. No. 4.4e-49;
  Matches  145;  Conservative   57;  Mismatches   90;  Indels    4;  Gaps    3;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAI-GVVPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              | |::||||: ||| : :  :: :|:: : |  ||||:: || | ||  |:  | || | :
Db          3 AGILNGKLVSAQIRDRLKNQVTRMQEQVPGFTPGLAILQVGDRDDSNLYINVKLKAAEEI 62

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPL--PHHMNDENILNAVSIEKD 117
              |||:  : ||  |:| ||||:: : | | |||| :|||||   : :| | ::||:: |||
Db         63 GIKATHIKLPRTSTESEVLKYVISLNEDASVHGFIVQLPLDSENSINTEAVINAIAPEKD 122

Qy        118 VDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAAL 177
              |||   :: |:||    :  |:||||||:||:   ||:| |: |||:||| ||| |
Db        123 VDGLTSVSAGKLARGDLNDCFIPCTPKGCLELIKEAGVQIAGRHAVVVGRSKIVGAPMHD 182

Qy        178 LLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDD 237
              ||   ||||:  || |   ::   : ||:: | |   :|:| ||||| :|| ||| | |
Db        183 LLLWNNATVTTCHSKTANLDKEVNKGDILVVATGQPEMVKGEWIKPGAVVIDCGINYVPD 242

Qy        238 PESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR-IHKFK 292
              . | | ::|||| |:|| : |  ||||||||||||:|||: :|:|||:| : |||
Db        243 DTKPNGRKVVGDVAYDEAKERASFITPVPGGVGPMTVAMLMQSTVESAQRFLQKFK 298


RESULT 4
C1TC_HUMAN
ID   C1TC_HUMAN     STANDARD;      PRT;   934 AA.
AC   P11586; Q86VC9;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   C-1-tetrahydrofolate synthase, cytoplasmic (C1-THF synthase)
DE   [Includes: Methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5);
DE   Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9);
DE   Formyltetrahydrofolate synthetase (EC 6.3.4.3)].
GN   MTHFD1 OR MTHFD OR MTHFC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-30.
RC   TISSUE=Liver;
RX   MEDLINE=89034046; PubMed=3053686;
RA   Hum D.W., Bell A.W., Rozen R., Mackenzie R.E.;
RT   "Primary structure of a human trifunctional enzyme. Isolation of a
RT   cDNA encoding methylenetetrahydrofolate dehydrogenase-
RT   methenyltetrahydrofolate cyclohydrolase-formyltetrahydrofolate
RT   synthetase.";
RL   J. Biol. Chem. 263:15946-15950(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Lymph;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 1-16.
RC   TISSUE=Platelet;
RX   MEDLINE=22608298; PubMed=12665801;
RA   Gevaert K., Goethals M., Martens L., Van Damme J., Staes A.,
RA   Thomas G.R., Vandekerckhove J.;
RT   "Exploring proteomes and analyzing protein processing by mass
RT   spectrometric identification of sorted N-terminal peptides.";
RL   Nat. Biotechnol. 21:566-569(2003).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 1-301.
RX   MEDLINE=98179934; PubMed=9519408;
RA   Allaire M., Li Y., Mackenzie R.E., Cygler M.;
RT   "The 3-D structure of a folate-dependent dehydrogenase/cyclohydrolase
RT   bifunctional enzyme at 1.5-A resolution.";
RL   Structure 6:173-182(1998).
RN   [5]
RP   VARIANT SPINA BIFIDA HIS-292, AND VARIANT GLN-652.
RX   MEDLINE=98272628; PubMed=9611072;
RA   Hol F.A., van der Put N.M.J., Geurds M.P.A., Heil S.G.,
RA   Trijbels F.J.M., Hamel B.C.J., Mariman E.C.M., Blom H.J.;
RT   "Molecular genetic analysis of the gene encoding the trifunctional
RT   enzyme MTHFD (methylenetetrahydrofolate-dehydrogenase,
RT   methenyltetrahydrofolate-cyclohydrolase, formyltetrahydrofolate
RT   synthetase) in patients with neural tube defects.";
RL   Clin. Genet. 53:119-125(1998).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +
CC       phosphate + 10-formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- DISEASE: Defects in MTHFD1 seem to be associated with elevated
CC       risk for neural tube defects (NTD) such as spina bifida.
CC   -!- MISCELLANEOUS: This trifunctional enzyme consists of two major
CC       domains: a N-terminal part, containing the methylene-THF
CC       dehydrogenase and the methenyl-THF cyclohydrolase activities and a
CC       larger formyl-THF synthetase domain.
```

```
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; TO OTHER DEHYDROGENASE/
CC       CYCLOHYDROLASE ENZYMES OR DOMAINS.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG TO OTHER
CC       FORMYLTETRAHYDROFOLATE SYNTHETASE DOMAINS OR ENZYMES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04031; AAA59574.1; -.
DR   EMBL; BC009806; AAH09806.1; -.
DR   EMBL; BC050420; AAH50420.1; -.
DR   PIR; A31903; A31903.
DR   PDB; 1A4I; 09-FEB-99.
DR   PDB; 1DIA; 09-OCT-02.
DR   PDB; 1DIB; 05-JUL-00.
DR   PDB; 1DIG; 05-JUL-00.
DR   SWISS-2DPAGE; P11586; HUMAN.
DR   Genew; HGNC:7432; MTHFD1.
DR   MIM; 172460; -.
DR   GO; GO:0005739; C:mitochondrion; TAS.
DR   GO; GO:0004329; F:formate-tetrahydrofolate ligase activity; TAS.
DR   GO; GO:0004477; F:methenyltetrahydrofolate cyclohydrolase act. . .; TAS.
DR   GO; GO:0004486; F:methylenetetrahydrofolate dehydrogenase act. . .; TAS.
DR   InterPro; IPR000559; Fmtethyd_synth.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF01268; FTHFS; 1.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00721; FTHFS_1; 1.
DR   PROSITE; PS00722; FTHFS_2; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Hydrolase; Ligase; Oxidoreductase; NADP; Multifunctional enzyme;
KW   One-carbon metabolism; ATP-binding; Purine biosynthesis;
KW   Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis; Disease mutation; Polymorphism; 3D-structure.
FT   INIT_MET      0      0
FT   DOMAIN        1    304       METHYLENETETRAHYDROFOLATE DEHYDROGENASE
FT                                AND CYCLOHYDROLASE.
FT   DOMAIN      305    934       FORMYLTETRAHYDROFOLATE SYNTHETASE.
FT   NP_BIND     379    386       ATP (BY SIMILARITY).
FT   VARIANT     133    133       R -> K (in dbSNP:1950902).
FT                                /FTId=VAR_016232.
FT   VARIANT     292    292       R -> H (in spina bifida).
FT                                /FTId=VAR_010241.
FT   VARIANT     652    652       R -> Q (in dbSNP:2236225).
FT                                /FTId=VAR_010251.
FT   CONFLICT    768    768       L -> F (IN REF. 2; AAH50420).
FT   STRAND        4      5
FT   HELIX         8     29
FT   TURN         31     32
FT   STRAND       36     42
FT   HELIX        46     62
FT   TURN         63     63
FT   STRAND       65     71
FT   TURN         73     74
FT   HELIX        77     89
FT   TURN         91     92
FT   STRAND       95     98
FT   HELIX       110    115
FT   TURN        116    116
FT   HELIX       119    121
FT   STRAND      122    122
FT   TURN        123    124
FT   HELIX       128    135
FT   TURN        136    137
FT   TURN        139    141
FT   HELIX       146    157
FT   TURN        158    158
FT   TURN        163    164
FT   STRAND      166    170
FT   TURN        174    176
FT   HELIX       177    186
FT   TURN        187    188
FT   STRAND      190    194
FT   TURN        196    197
FT   HELIX       201    205
FT   TURN        206    207
FT   STRAND      210    213
FT   TURN        218    219
FT   STRAND      221    221
FT   HELIX       223    225
FT   TURN        228    229
FT   STRAND      231    234
FT   STRAND      238    241
FT   TURN        243    244
FT   TURN        246    247
FT   STRAND      249    252
FT   STRAND      255    255
FT   HELIX       257    260
FT   TURN        261    263
FT   STRAND      266    267
FT   TURN        270    272
FT   TURN        275    276
FT   HELIX       277    294
SQ   SEQUENCE   934 AA;  101428 MW;  4AA4678A2BA065E4 CRC64;

  Query Match             49.5%;  Score 735;  DB 1;  Length 934;
  Best Local Similarity   49.0%;  Pred. No. 7.5e-49;
  Matches  145;  Conservative   56;  Mismatches   91;  Indels    4;  Gaps    3;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAI-GVVPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              |:|::|| :: ||| . :  :: ::|: : |  | ||:: ||:| ||  |:  | || | :
Db          3 AEILNGKEISAQIRARLKNQVTQLKEQVPGFTPRLAILQVGNRDDSNLYINVKLKAAEEI 62

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPL--PHHMNDENILNAVSIEKD 117
              |||:  : ||  ::| ||:|:| : | | :||| ||||||   : :| | ::||:: |||
Db         63 GIKATHIKLPRTTTESEVMKYITSLNEDSTVHGFLVQLPLDSENSINTEEVINAIAPEKD 122

Qy        118 VDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAAL 177
              |||   :| ||||    :  |:|||||||:||:   || | |: |||:||| ||| |
Db        123 VDGLTSINAGRLARGDLNDCFIPCTPKGCLELIKETGVPIAGRHAVVVGRSKIVGAPMHD 182

Qy        178 LLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDD 237
              ||   ||||:  || |   :|   : ||:: | |   :|:| |||||| :|| ||| | |
Db        183 LLLWNNATVTTCHSKTAHLDEEVNKGDILVVATGQPEMVKGEWIKPGAIVIDCGINYVPD 242

Qy        238 PESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR-IHKFK 292
              : | | ::|||| |:|| : |  |||||||||||:|||: :|:|||| : |||
Db        243 DKKPNGRKVVGDVAYDEAKERASFITPVPGGVGPMTVAMLMQSTVESAKRFLEKFK 298


RESULT 5
C1TC_RAT
ID   C1TC_RAT       STANDARD;      PRT;   934 AA.
AC   P27653; Q62808;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   C-1-tetrahydrofolate synthase, cytoplasmic (C1-THF synthase)
DE   [Includes: Methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5);
DE   Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9);
DE   Formyltetrahydrofolate synthetase (EC 6.3.4.3)].
GN   MTHFD1 OR MTHFD.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90243660; PubMed=2186031;
RA   Thigpen A.E., West M.G., Appling D.R.;
RT   "Rat C1-tetrahydrofolate synthase. cDNA isolation, tissue-specific
RT   levels of the mRNA, and expression of the protein in yeast.";
RL   J. Biol. Chem. 265:7907-7913(1990).
RN   [2]
RP   REVISIONS TO 18 AND 23.
RC   TISSUE=Liver;
RA   Appling D.R.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +
CC       phosphate + 10-formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- MISCELLANEOUS: This trifunctional enzyme consists of two major
CC       domains: a N-terminal part, containing the methylene-THF
CC       dehydrogenase and the methenyl-THF cyclohydrolase activities and a
CC       larger formyl-THF synthetase domain.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; TO OTHER DEHYDROGENASE/
CC       CYCLOHYDROLASE ENZYMES OR DOMAINS.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG TO OTHER
CC       FORMYLTETRAHYDROFOLATE SYNTHETASE DOMAINS OR ENZYMES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J05519; AAA74248.1; -.
DR   EMBL; U31032; AAA74744.1; -.
DR   HSSP; P11586; 1A4I.
DR   InterPro; IPR000559; Fmtethyd_synth.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF01268; FTHFS; 1.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00721; FTHFS_1; 1.
DR   PROSITE; PS00722; FTHFS_2; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Hydrolase; Ligase; Oxidoreductase; NADP; Multifunctional enzyme;
KW   One-carbon metabolism; ATP-binding; Purine biosynthesis;
KW   Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    304       METHYLENETETRAHYDROFOLATE DEHYDROGENASE
FT                                AND CYCLOHYDROLASE.
FT   DOMAIN      305    934       FORMYLTETRAHYDROFOLATE SYNTHETASE.
FT   NP_BIND     379    386       ATP (BY SIMILARITY).
SQ   SEQUENCE   934 AA;  100864 MW;  0D48DEE28F9648FF CRC64;

  Query Match             49.2%;  Score 731;  DB 1;  Length 934;
  Best Local Similarity   48.3%;  Pred. No. 1.5e-48;
  Matches  143;  Conservative   57;  Mismatches   92;  Indels    4;  Gaps    3;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAI-GVVPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              | |::||:|: |||   :  :: :|:: : |  ||||:: || | ||  |:  | || : :
Db          3 AGILNGKVVSAQIRNRLKTQVTQMQEQVPGFTPGLAILQVGDRDDSNLYINVKLKAAQEI 62

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPL--PHHMNDENILNAVSIEKD 117
                |||:  : ||  |:| ||||:: : | | :||| :|||||   : :| | ::||:: |||

Db         63 GIKATHIKLPRTSTESEVLKYVISLNEDATVHGFIVQLPLDSENSINTEAVINAIAPEKD 122

Qy        118 VDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAAL 177
                |||   :| |:||       |:|||||:||:   ||:| |: |||:||| ||| |
Db        123 VDGLTSINAGKLARGDLKDCFIPCTPKGCLELIKETGVQIAGRHAVVVGRSKIVGAPMHD 182

Qy        178 LLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDD 237
              ||   ||||:  || |   ::   : ||:: | |   :|:| |||||| :|| ||| | |
Db        183 LLLWNNATVTTCHSKTADLDKEVNKGDILVVATGQPEMVKGEWIKPGAVVIDCGINYVPD 242

Qy        238 PESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR-IHKFK 292
                | | ::|||| |:|| : |  ||||||||||||:|||: :|:|||:| : |||
Db        243 DTKPNGRKVVGDVAYDEAKEKASFITPVPGGVGPMTVAMLMQSTVESAQRFLKKFK 298


RESULT 6
C1TM_YEAST
ID   C1TM_YEAST     STANDARD;      PRT;   975 AA.
AC   P09440;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   C-1-tetrahydrofolate synthase, mitochondrial precursor (C1-THF
DE   synthase) [Includes: Methylenetetrahydrofolate dehydrogenase
DE   (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9);
DE   Formyltetrahydrofolate synthetase (EC 6.3.4.3)].
GN   MIS1 OR YBR084W OR YBR0751.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88227973; PubMed=2836393;
RA   Shannon K.W., Rabinowitz J.C.;
RT   "Isolation and characterization of the Saccharomyces cerevisiae MIS1
RT   gene encoding mitochondrial C1-tetrahydrofolate synthase.";
RL   J. Biol. Chem. 263:7717-7725(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RA   Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
RA   Vissers S.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +
CC       phosphate + 10-formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- MISCELLANEOUS: THIS TRIFUNCTIONAL ENZYME CONSISTS OF TWO MAJOR
CC       DOMAINS: AN AMINOTERMINAL PART, CONTAINING THE METHYLENE-THF
CC       DEHYDROGENASE AND THE METHENYL-THF CYCLOHYDROLASE ACTIVITIES AND A
CC       LARGER FORMYL-THF SYNTHETASE DOMAIN.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; TO OTHER DEHYDROGENASE/
CC       CYCLOHYDROLASE ENZYMES OR DOMAINS.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG TO OTHER
CC       FORMYLTETRAHYDROFOLATE SYNTHETASE DOMAINS OR ENZYMES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35953; CAA85029.1; -.
DR   EMBL; J03724; AAA34781.1; -.
DR   PIR; A28174; A28174.
```

```
DR   HSSP; P11586; 1A4I.
DR   GermOnline; 138627; -.
DR   SGD; S0000288; MIS1.
DR   GO; GO:0005739; C:mitochondrion; IDA.
DR   GO; GO:0004329; F:formate-tetrahydrofolate ligase activity; IDA.
DR   GO; GO:0004477; F:methenyltetrahydrofolate cyclohydrolase act. . .; IDA.
DR   GO; GO:0004488; F:methylenetetrahydrofolate dehydrogenase (NA. . .; IDA.
DR   GO; GO:0001718; P:conversion of met-tRNAf to fmet-tRNA; IMP.
DR   GO; GO:0006139; P:nucleobase, nucleoside, nucleotide and nucl. . .; IMP.
DR   InterPro; IPR000559; Fmtethyd_synth.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF01268; FTHFS; 1.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00721; FTHFS_1; 1.
DR   PROSITE; PS00722; FTHFS_2; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Hydrolase; Ligase; Oxidoreductase; NADP; Multifunctional enzyme;
KW   Mitochondrion; Transit peptide; One-carbon metabolism; ATP-binding;
KW   Purine biosynthesis; Amino-acid biosynthesis; Histidine biosynthesis;
KW   Methionine biosynthesis.
FT   TRANSIT       1     34       MITOCHONDRION.
FT   CHAIN        35    975       C-1-TETRAHYDROFOLATE SYNTHASE.
FT   DOMAIN       35    343       METHYLENETETRAHYDROFOLATE DEHYDROGENASE
FT                                AND CYCLOHYDROLASE.
FT   DOMAIN      344    975       FORMYLTETRAHYDROFOLATE SYNTHETASE.
FT   NP_BIND     408    415       ATP (BY SIMILARITY).
SQ   SEQUENCE   975 AA;  106217 MW;  B9421D4F8981531F CRC64;

  Query Match             48.0%;  Score 713.5;  DB 1;  Length 975;
  Best Local Similarity   50.7%;  Pred. No. 3.6e-47;
  Matches  145;  Conservative   45;  Mismatches   95;  Indels    1;  Gaps    1;

Qy          3 IIDGKLVAKQIREEIAVEIAKMK-DAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              |: |: :|: |||:   ||  :|       | | :| ||:| || |||| | || :   :
Db         37 ILSGRKLAQSIREKANDEIQAIKLKHPNFKPTLKIIQVGARPDSSTYVRMKLKASKDSNV 96

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
                 || : :| |:|| |:  | | |:||:|:||||| |::: | ||| :||||||
Db         97 DCIIEKLPAEITEVELLKKISDINDDDSIHGLLIQLPLPRHLDETTITNAVDFKKDVDGF 156

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              |  | | || :|  |:|:|||| |||:||    |:: || |||:|||:||| | | ||:
Db        157 HRYNAGELAKKGGKPYFIPCTPYGCMKLLEEAHVKLDGKNAVVLGRSSIVGNPIASLLKN 216

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              |||||:: ||:|:|   |:  |||||||| |:   |:  ||| || :|||||  | |
Db        217 ANATVTVCHSHTRNIAEVVSQADIVIAACGIPQYVKSDWIKEGAVVIDVGINYVPDISKK 276

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
               | :|||| ::   :    |||||||||||:|||:|| | :|||
Db        277 SGQKLVGDVDFDSVKEKTSYITPVPGGVGPMTVAMLVSNVLLAAKR 322


RESULT 7
FOLD_STRTR
ID   FOLD_STRTR     STANDARD;      PRT;   284 AA.
AC   P96050;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   FolD bifunctional protein [Includes: Methylenetetrahydrofolate
DE   dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase
DE   (EC 3.5.4.9)].
GN   FOLD.
OS   Streptococcus thermophilus.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1308;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sfi6;
RX   MEDLINE=97084576; PubMed=8930919;
RA   Stingele F., Mollet B.;
RT   "Disruption of the gene encoding penicillin-binding protein 2b
RT   (pbp2b) causes altered cell morphology and cease in exopolysaccharide
RT   production in Streptococcus thermophilus Sfi6.";
RL   Mol. Microbiol. 22:357-366(1996).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC       DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U58210; AAC44612.1; -.
DR   HSSP; P11586; 1A4I.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
KW   Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis; Hydrolase.
SQ   SEQUENCE   284 AA;  31383 MW;  B23ED2124377AF88 CRC64;

  Query Match             46.6%;  Score 692;  DB 1;  Length 284;
  Best Local Similarity   49.1%;  Pred. No. 3.6e-46;
  Matches  141;  Conservative   53;  Mismatches   85;  Indels    8;  Gaps    4;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | |:||| :|  ::|::  ::|::|:   :|||| ||:||    || |||||::| :  |
Db          2 AIIMDGKALAVNMQEQLQEKVARLKEKEWIVPGLVVIMVGENPASQVYVRNKERAAKKAG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
                |  ||| |  ||:|::: |  :| :|  ||||||||||:|:|:  || |:  :|||||
Db         62 FHSKTVNLSESISEEELIEVIEKYNQNPLFHGILVQLPLPNHINEMRILLAIDPKKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||:| | |   || |  ||||| | ||:|  | ||::|| ||:|||||||| | | ||
Db        122 FHPMNTGNL-WNGR-PQMVPCTPAGIMEILREYNVELEGKTAVIIGRSNIVGKPMAQLLL 179

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              : ||||:: || |    ::  :||::| |:| |  |   ::| || :||||  | |
Db        180 EKNATVTLTHSRTPHLAKVCNKADVLIVAIGRAKFVTEEFVKEGAVVIDVGIN--RDEEG 237

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                :| ||| :::  :   :||||||||||||| ||:  | ::| |
Db        238 ----KLCGDVDFDQVKEKVSMITPVPGGVGPMTITMLMEQTYQAALR 280


RESULT 8
FOLD_BACSU
ID   FOLD_BACSU     STANDARD;      PRT;   283 AA.
AC   P54382;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   FolD bifunctional protein [Includes: Methylenetetrahydrofolate
```

```
DE   dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase
DE   (EC 3.5.4.9)].
GN   FOLD OR BSU24310.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE=97124195; PubMed=8969508;
RA   Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA   Kobayashi Y.;
RT   "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT   the Bacillus subtilis genome containing the skin element and many
RT   sporulation genes.";
RL   Microbiology 142:3103-3111(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC       DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D84432; BAA12572.1; -.
DR   EMBL; Z99116; CAB14362.1; -.
DR   PIR; E69626; E69626.
DR   HSSP; P11586; 1A4I.
DR   SubtiList; BG11711; folD.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
KW   Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis; Hydrolase; Complete proteome.
SQ   SEQUENCE   283 AA;  30685 MW;  E004E0AC40A30889 CRC64;

  Query Match             46.0%;  Score 683.5;  DB 1;  Length 283;
  Best Local Similarity   49.1%;  Pred. No. 1.6e-45;
  Matches  141;  Conservative   50;  Mismatches   87;  Indels    9;  Gaps    4;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | ||||  |:: ||::| |: ::|    || |||||||:|     | :||| |||| | :|
Db          3 ATIIDGKETAREKREQLAKEVEELKKQ-GVTPGLAVILIGDDPASHSYVRGKKKAAETMG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              :          :| |:|  |  :| ||  |||||||||| |:::: ::  :| :|||||
Db         62 MNFKLDQFDSSLTEAELLSIIDQYNQDPEFHGILVQLPLPDHISEKAVIERISPDKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
             |||||:|:: : | |  |:|||| | :||| :  ::: ||  ||:|||||| |   ||
Db        122 FHPLNVGKMLL-GEDT-PLPCTPHGIVELLKKTNIDLSGKEVVVVGRSNIVGKPVGQLLL 179

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||||:  || |:   | |::|||:: ||| || :    || || :||||:| :::
Db        180 NENATVTYCHSRTENITEHTKKADILVVAVGRANFISADQIKEGAVVIDVGVNRLENG-- 237

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                :| ||| :| | : |  |||||||||||| ||  ||::||||
Db        238 ----KLCGDVEFEGAKEKASFITPVPGGVGPMTITMLAHNTVKSAKR 280


RESULT 9
FOLD_HAEIN
ID   FOLD_HAEIN     STANDARD;      PRT;   281 AA.
AC   P44313;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   FolD bifunctional protein [Includes: Methylenetetrahydrofolate
DE   dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase
DE   (EC 3.5.4.9)].
GN   FOLD OR HI0609.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
```

```
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC       DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32743; AAC22268.1; -.
DR   PIR; A64081; A64081.
DR   HSSP; P11586; 1A4I.
DR   TIGR; HI0609; -.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
KW   Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis; Hydrolase; Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   281 AA;  30329 MW;  952ABAB9778CF51C CRC64;

  Query Match             45.9%;  Score 681.5;  DB 1;  Length 281;
  Best Local Similarity   51.1%;  Pred. No. 2.3e-45;
  Matches  144;  Conservative   44;  Mismatches   85;  Indels    9;  Gaps    2;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |||| |  ::|||:  :| :|     :     |||||||||:   || || ||:|:|| ||
Db          2 AKIISGTELSKQIKANLADKITHYIEQGKRAPGLAVILVGADPASQIYVGNKRKSCEEVG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | |   :||| ::::|:|  |    |:| :: |||||||||  :| | |:  :  :|||||
Db         62 ILSKSYDLPETTTQNELLAIIDQLNADKNIDGILVQLPLPKQINAEAIIERIDPKKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              ||| |:|||    | |   .||| | |:||   |::: || ||::| ||||| | :| |
Db        122 FHPYNVGRLCQ--RIPTLRACTPYGVMKLLETTGIDLHGKHAVIVGASNIVGRPMSLELL 179

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              | |||:: |  ||  |   |||||:: |||  ||: | |||  | :|||||| ||
Db        180 LAGATVTVTHRFTKNLENHVRQADILVVAVGKPNLISGDWIKESAVVIDVGINRVDG--- 236

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTL 282
                :||||| :::|:: |  |||||||||||:|||:||||
Db        237 ----KLVGDVEFDKAAEKAAYITPVPGGVGPMTVAMLMSNTL 274


RESULT 10
FOLD_ECOLI
ID   FOLD_ECOLI     STANDARD;      PRT;   287 AA.
AC   P24186; P77132;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   FolD bifunctional protein [Includes: Methylenetetrahydrofolate
DE   dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase
DE   (EC 3.5.4.9)].
GN   FOLD OR ADS OR B0529.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-35.
RX   MEDLINE=92084696; PubMed=1748668;
RA   D'Ari L., Rabinowitz J.C.;
RT   "Purification, characterization, cloning, and amino acid sequence of
RT   the bifunctional enzyme 5,10-methylenetetrahydrofolate
RT   dehydrogenase/5,10-methenyltetrahydrofolate cyclohydrolase from
RT   Escherichia coli.";
RL   J. Biol. Chem. 266:23953-23958(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RA   Yonetani Y., Sanpei G., Mizobuchi K.;
RL   Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA   Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RT   "Sequence of minutes 4-25 of Escherichia coli.";
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer.
CC   -!- MISCELLANEOUS: THIS ENZYME IS SPECIFIC FOR NADP.
CC   -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC       DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M74789; AAA23803.1; -.
DR   EMBL; D10588; BAA01445.1; -.
DR   EMBL; AE000159; AAC73631.1; -.
DR   EMBL; U82664; AAB40282.1; -.
DR   PIR; H64784; JS0662.
DR   PDB; 1B0A; 29-JUN-99.
DR   EcoGene; EG10328; folD.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
KW   Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis; Hydrolase; Complete proteome; 3D-structure.
FT   INIT_MET      0      0
FT   CONFLICT     46     46       S -> L (IN REF. 4).
FT   CONFLICT    199    199       V -> L (IN REF. 1).
SQ   SEQUENCE   287 AA;  30912 MW;  163F6BDD9024514E CRC64;

  Query Match             45.7%;  Score 679;  DB 1;  Length 287;
  Best Local Similarity   48.6%;  Pred. No. 3.6e-45;
  Matches  141;  Conservative   50;  Mismatches   89;  Indels   10;  Gaps    4;
```

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGV-VPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              ||||||| :|:|:| |:| :: : : | |:  |||||:||||    || || :|:|||| |
Db          2 AKIIDGKTIAQQVRSEVAQKV-QARIAAGLRAPGLAVVLVGSNPASQIYVASKRKACEEV 60

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVD 119
               |  |   :||| :|| |:|: | | |:| :: |||||||||  :::  :|  :  :||||
Db         61 GFVSRSYDLPETTSEAELLELIDTLNADNTIDGILVQLPLPAGIDNVKVLERIHPDKDVD 120

Qy        120 GFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLL 179
              |||| |:|||    | |   ||||:| : || || ::  |  ||||| ||||| | :: |
Db        121 GFHPYNVGRLCQ--RAPRLRPCTPRGIVTLLERYNIDTFGLNAVVIGASNIVGRPMSMEL 178

Qy        180 QKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPE 239
                |  | :: |  ||         ||::| |||    : | ||| || :|||||| :::
Db        179 LLAGCTTTVTHRFTKNLRHHVENADLLIVAVGKPGFIPGDWIKEGAIVIDVGINRLENG- 237

Qy        240 SPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIH 289
                ::|||| :|:|:| |  ||||||||||||:| |: |||::    |
Db        238 -----KVVGDVVFEDAAKRASYITPVPGGVGPMTVATLIENTLQACVEYH 282


RESULT 11
FOLD_PHOPO
ID   FOLD_PHOPO     STANDARD;      PRT;   285 AA.
AC   P51696;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   FolD bifunctional protein [Includes: Methylenetetrahydrofolate
DE   dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase
DE   (EC 3.5.4.9)].
GN   FOLD.
OS   Photobacterium phosphoreum.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Photobacterium.
OX   NCBI_TaxID=659;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIMB 844;
RA   Pawelek P.D., Meija N., Mackenzie R.E.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC       DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34207; AAA98507.1; -.
DR   PIR; S71924; S71924.
DR   HSSP; P11586; 1A4I.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
KW   Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis; Hydrolase.

SQ   SEQUENCE   285 AA;  30775 MW;  9EB1FC954EAA4396 CRC64;

  Query Match             45.5%;  Score 676;  DB 1;  Length 285;
  Best Local Similarity   48.8%;  Pred. No. 6.1e-45;
  Matches  141;  Conservative   46;  Mismatches   94;  Indels    8;  Gaps    2;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |:|||||:::: :|:|:|  :    ||    |||||:|||    || || :|:|||| ||
Db          3 AQIIDGKIISQTVRQEVAARVKARTDAGLRAPGLAVVLVGQDPASQIYVGSKRKACEEVG 62

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
               |   :||  :|| ::|  |   | ||:: ||||||||  |:   ||  :  ||||||
Db         63 FISKSFDLPSSASEQQLLDLIDELNQDPTMDGILVQLPLPAGMDCTRILERIDPEKDVDG 122

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              ||| |:|||:   | |    |||||  : || || :|::|| ||::| |||||  |  | |
Db        123 FHPYNVGRLSQ--RIPKLRSCTPKGIITLLERYNIEVRGKHAVIVGASNIVGRPMTLELL 180

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              | || :  |  |:  |   |||||:: |||  | : | ||| || :|||||| :::
Db        181 LAGATTTTCHRFTQDLEGHIRQADILVVAVGKPNFIPGGWIKEGATVIDVGINRLENG-- 238

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIH 289
                :| ||| :: | : |  ||||||||||||:| |: ||| : :: |
Db        239 ----KLCGDVEFDVACQRAKYITPVPGGVGPMTVASLIENTLLACEQYH 283


RESULT 12
FOLD_SALTI
ID   FOLD_SALTI     STANDARD;      PRT;   287 AA.
AC   Q60006;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   FolD bifunctional protein [Includes: Methylenetetrahydrofolate
DE   dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase
DE   (EC 3.5.4.9)].
GN   FOLD OR STY0588 OR T2321.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STY4;
RX   MEDLINE=94327079; PubMed=8050713;
RA   Rossolini G.M., Muscas P., Chiesurin A., Satta G.;
RT   "fimA-folD genes linkage in Salmonella identifies a putative
RT   junctional site of chromosomal rearrangement in the enterobacterial
RT   genome.";
RL   FEMS Microbiol. Lett. 119:321-328(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
```

```
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- MISCELLANEOUS: THIS ENZYME IS SPECIFIC FOR NADP (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC       DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74603; CAA52683.1; -.
DR   EMBL; AL627267; CAD05024.1; -.
DR   EMBL; AE016841; AAO69915.1; -.
DR   PIR; S36633; S36633.
DR   HSSP; P11586; 1A4I.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
KW   Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis; Hydrolase; Complete proteome.
FT   INIT_MET      0       0       BY SIMILARITY.
SQ   SEQUENCE   287 AA;  30693 MW;  DA4E18D51701B14D CRC64;

  Query Match             45.3%;  Score 673;  DB 1;  Length 287;
  Best Local Similarity   47.9%;  Pred. No. 1e-44;
  Matches  139;  Conservative   52;  Mismatches   89;  Indels   10;  Gaps    4;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGV-VPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              ||||||| :|:|:| |:| : : | |:  |||||:||||   || || :|:|||: |
Db          2 AKIIDGKTIAQQVRSEVAQKV-QARVAAGLRAPGLAVVLVGSNPASQIYVASKRKACDEV 60

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVD 119
              |  |:  :||| :|| |:|  | | |:| :: |||||||||  :::  :|  :: :||||
Db         61 GFVSHSYDLPETTSEAELLALIDTLNADNTIDGILVQLPLPAGIDNVKVLERIAPDKDVD 120

Qy        120 GFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLL 179
              |||| |:|||    | |   ||||:| : || || ::  |  ||||| ||||| | :: |
Db        121 GFHPYNVGRLCQ--RAPRLRPCTPRGIVTLLERYNIDTYGLNAVVIGASNIVGRPMSMEL 178

Qy        180 QKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPE 239
                | | :: |  ||         ||::| |||    : | ||| || :|||||| :::
Db        179 LLAGCTTTVTHRFTKDLRHHVEHADLLIVAVGKPGFIPGEWIKEGAIVIDVGINRLENG- 237

Qy        240 SPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIH 289
                ::|||| ::||:  |  ||||||||||||:| |: |||::    |
Db        238 -----KVVGDVVFDEAAARASYITPVPGGVGPMTVATLIENTLQACIEYH 282


RESULT 13
FOLD_SALTY
ID   FOLD_SALTY     STANDARD;      PRT;   287 AA.
AC   P58688;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   FolD bifunctional protein [Includes: Methylenetetrahydrofolate
DE   dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase
DE   (EC 3.5.4.9)].
GN   FOLD OR STM0542.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC       5,10-methenyltetrahydrofolate + NADPH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- PATHWAY: Necessary for the biosynthesis of purines, thymydylate,
CC       methionine, histidine, pantothenate, and formyl tRNA-Met.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- MISCELLANEOUS: THIS ENZYME IS SPECIFIC FOR NADP (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC       DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008721; AAL19496.1; -.
DR   StyGene; SG?????; foID.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
KW   Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
KW   Histidine biosynthesis; Hydrolase; Complete proteome.
FT   INIT_MET      0       0       BY SIMILARITY.
SQ   SEQUENCE   287 AA;  30712 MW;  26A97583C26CE103 CRC64;

  Query Match             45.0%;  Score 669;  DB 1;  Length 287;
  Best Local Similarity   47.9%;  Pred. No. 2.1e-44;
  Matches  139;  Conservative   51;  Mismatches   90;  Indels   10;  Gaps    4;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGV-VPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              ||||||| :|:|:| |:| : : | |:  |||||:||||   || || :|:|||: |
Db          2 AKIIDGKTIAQQVRSEVAQKV-QARVAAGLRAPGLAVVLVGSNPASQIYVASKRKACDEV 60

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVD 119
              |  |   :||| :|| |:|  | | |:| :: |||||||||  :::  :|  :: :||||
Db         61 GFVSRSYDLPETTSEAELLALIDTLNADNTIDGILVQLPLPAGIDNVKVLERIAPDKDVD 120

Qy        120 GFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLL 179
              |||| |:|||    | |   ||||:| : || || ::  |  ||||| ||||| | :: |
Db        121 GFHPYNVGRLCQ--RAPRLRPCTPRGIVTLLERYNIDTYGLNAVVIGASNIVGRPMSMEL 178

Qy        180 QKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPE 239
                | | :: |  ||         ||::| |||    : | ||| || :|||||| :::
Db        179 LLAGCTTTVTHRFTKDLRHHVEHADLLIVAVGKPGFIPGEWIKEGAIVIDVGINRLENG- 237

Qy        240 SPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIH 289
```

```
                    ::|||| ::||:  |  ||||||||||||:| |: |||::    |
Db        238 -----KVVGDVVFDEAAARASYITPVPGGVGPMTVATLIENTLQACIEYH 282


RESULT 14
MTDC_DROME
ID   MTDC_DROME     STANDARD;      PRT;   357 AA.
AC   Q04448;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase,
DE   mitochondrial precursor [Includes: NAD-dependent
DE   methylenetetrahydrofolate dehydrogenase (EC 1.5.1.15);
DE   Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9)].
GN   NMDMC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93250056; PubMed=8485162;
RA   Price B.D., Laughon A.;
RT   "The isolation and characterization of a Drosophila gene encoding a
RT   putative NAD-dependent methylenetetrahydrofolate dehydrogenase-
RT   methenyltetrahydrofolate cyclohydrolase.";
RL   Biochim. Biophys. Acta 1173:94-98(1993).
CC   -!- FUNCTION: MAY PLAY SOME ROLE IN SPERMATOGENESIS.
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NAD(+) =
CC       5,10-methenyltetrahydrofolate + NADH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- COFACTOR: Magnesium. This NAD-dependent bifunctional enzyme has
CC       very different kinetic properties than the larger NADP-dependent
CC       trifunctional enzyme and is unique in that it requires formation
CC       of an enzyme-magnesium complex to allow binding of NAD.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN DEVELOPING TISSUES AND AT HIGH
CC       LEVELS IN ADULT TISSUES.
CC   -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC       DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07958; AAB41352.1; -.
DR   PIR; S32562; S32562.
DR   HSSP; P11586; 1A4I.
DR   FlyBase; FBgn0010222; Nmdmc.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NAD;
KW   Hydrolase; Mitochondrion; Transit peptide; Magnesium.
FT   TRANSIT       1     53       MITOCHONDRION (POTENTIAL).
FT   CHAIN        54    357       BIFUNCTIONAL METHYLENETETRAHYDROFOLATE
FT                                 DEHYDROGENASE/CYCLOHYDROLASE.
SQ   SEQUENCE   357 AA;  38759 MW;  ABD8F5D1119CC198 CRC64;

  Query Match             43.9%;  Score 652.5;  DB 1;  Length 357;
  Best Local Similarity   46.6%;  Pred. No. 5.2e-43;

  Matches  139;  Conservative   54;  Mismatches   90;  Indels   15;  Gaps    5;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |:||||| :|:::| ::| |:  |: |    | |  ::||    |: || ||  ||  ||
Db         56 AQIIDGKAIAQEVRTQLAHELKGMEAAGYPKPHLTAVIVGEDPASEKYVANKMVACREVG 115

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | |    ||  ::::|:|: ||  | || | |||||||:| |:|:  | ||| ::|||||
Db        116 ISSETKRLPASTTQEELLQLIADLNKDPQVTGILVQLPVPEHINERTICNAVDVDKDVDG 175

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLL- 179
              |: :|||| |:       :| || |   ||    :|  |: |||:||  | :| |:||
Db        176 FNEVNIGRTALDMEAN--IPATPLGVKRLLEHMKIETLGRNAVVVGRSKNVSLPMAILLH 233

Qy        180 -------QKANATVSIVHSNTKKPEEIT---RQADIVIAAVGVANLVRGSWIKPGAAIID 229
                    :  :|||:| |  |  |:|:    |||||:: |||   |:    :|||| :||
Db        234 ADGKYATKAMDATVTICHRYT-PPKELARHCRQADIIVVAVGKPGLITKDMVKPGACVID 292

Qy        230 VGINPVDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
              |||| : | ||   ::|||| :|| ::|| ||||||||||||:|||: |||::|::
Db        293 VGINRIKD-ESTGQFKLVGDVDFEEVRQVAGHITPVPGGVGPMTVAMLMHNTLKAARK 349


RESULT 15
MTDC_MOUSE
ID   MTDC_MOUSE     STANDARD;      PRT;   350 AA.
AC   P18155;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase,
DE   mitochondrial precursor [Includes: NAD-dependent
DE   methylenetetrahydrofolate dehydrogenase (EC 1.5.1.15);
DE   Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9)].
GN   MTHFD2 OR NMDMC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 36-55.
RX   MEDLINE=89174757; PubMed=2647744;
RA   Belanger C., Mackenzie R.E.;
RT   "Isolation and characterization of cDNA clones encoding the murine
RT   NAD-dependent methylenetetrahydrofolate dehydrogenase-
RT   methenyltetrahydrofolate cyclohydrolase.";
RL   J. Biol. Chem. 264:4837-4843(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91153660; PubMed=1999293;
RA   Belanger C., Mackenzie R.E.;
RT   "Structural organization of the murine gene encoding NAD-dependent
RT   methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate
RT   cyclohydrolase.";
RL   Gene 97:283-288(1991).
RN   [3]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=91360330; PubMed=1843253;
RA   Belanger C., Peri K.G., Mackenzie R.E.;
RT   "Analysis of the promoter region of the gene encoding NAD-dependent
RT   methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate
RT   cyclohydrolase.";
RL   Nucleic Acids Res. 19:4341-4345(1991).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NAD(+) =
CC       5,10-methenyltetrahydrofolate + NADH.
CC   -!- CATALYTIC ACTIVITY: 5,10-methenyltetrahydrofolate + H(2)O = 10-
CC       formyltetrahydrofolate.
CC   -!- COFACTOR: Magnesium. This NAD-dependent bifunctional enzyme has
CC       very different kinetic properties than the larger NADP-dependent
CC       trifunctional enzyme and is unique in that it requires formation
CC       of an enzyme-magnesium complex to allow binding of NAD.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
```

```
CC   -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC       DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04627; AAA39827.1; -.
DR   EMBL; M63445; AAA39828.1; -.
DR   EMBL; M63415; AAA39828.1; JOINED.
DR   EMBL; M63439; AAA39828.1; JOINED.
DR   EMBL; M63440; AAA39828.1; JOINED.
DR   EMBL; M63441; AAA39828.1; JOINED.
DR   EMBL; M63442; AAA39828.1; JOINED.
DR   EMBL; M63443; AAA39828.1; JOINED.
DR   EMBL; M63444; AAA39828.1; JOINED.
DR   EMBL; S52980; -; NOT_ANNOTATED_CDS.
DR   PIR; A33267; A33267.
DR   HSSP; P11586; 1A4I.
DR   MGD; MGI:1338850; Mthfd2.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NAD;
KW   Hydrolase; Mitochondrion; Transit peptide; Magnesium.
FT   TRANSIT       1     35       MITOCHONDRION.
FT   CHAIN        36    350       BIFUNCTIONAL METHYLENETETRAHYDROFOLATE
FT                                 DEHYDROGENASE/CYCLOHYDROLASE.
SQ   SEQUENCE   350 AA;  37863 MW;  896AD40D9154E9D7 CRC64;

  Query Match             42.4%;  Score 629.5;  DB 1;  Length 350;
  Best Local Similarity   46.2%;  Pred. No. 3e-41;
  Matches  138;  Conservative   52;  Mismatches   94;  Indels   15;  Gaps    5;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :| |: :|:||::|:  |: :   :    | |:|||||    | :|| || :|   ||
Db         37 AVVISGRKLAQQIKQEVQQEVEEWVASGNKRPHLSVILVGDNPASHSYVLNKTRAAAEVG 96

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | |  :  |   ||:|:|  |   |:| :| |:|||||| |::  : |||| :|||||
Db         97 INSETIVKPASVSEEELLNSIRKLNNDENVDGLLVQLPLPEHIDERKVCNAVSPDKDVDG 156

Qy        121 FHPLNIGRLAMQGRDPF-FVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLL 179
              || :|:||: :   | :  :| || |  |:: | |:   ||  || |||  |||| |:||
Db        157 FHVINVGRMCL---DQYSMLPATPWGVWEIIKRTGIPTLGKNVVVAGRSKNVGMPIAMLL 213

Qy        180 QK--------ANATVSIVHSNTKKPE--EITRQADIVIAAVGVANLVRGSWIKPGAAIID 229
                       :|||:| |  | | :  : |  |||||:| |: ||:    || |||:||
Db        214 HTDGAHERPGGDATVTISHRYTPKEQLKKHTILADIVISAAGIPNLITADMIKEGAAVID 273

Qy        230 VGINPVDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||| | || : :  :|||||| :|   | || ||||||||||||:|||: ||: :||::
Db        274 VGINRVQDPVTAKP-KLVGDVDFEGVKKKAGYITPVPGGVGPMTVAMLMKNTIIAAKKV 331


Search completed: July 24, 2004, 06:37:19
Job time : 54 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 24, 2004, 06:47:07 ; Search time 3406 Seconds
(without alignments)
2560.116 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10723061/runat_21072004_090401_7121/app_query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10723061_@CGN_1_1_6425_@runat_21072004_090401_7121 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

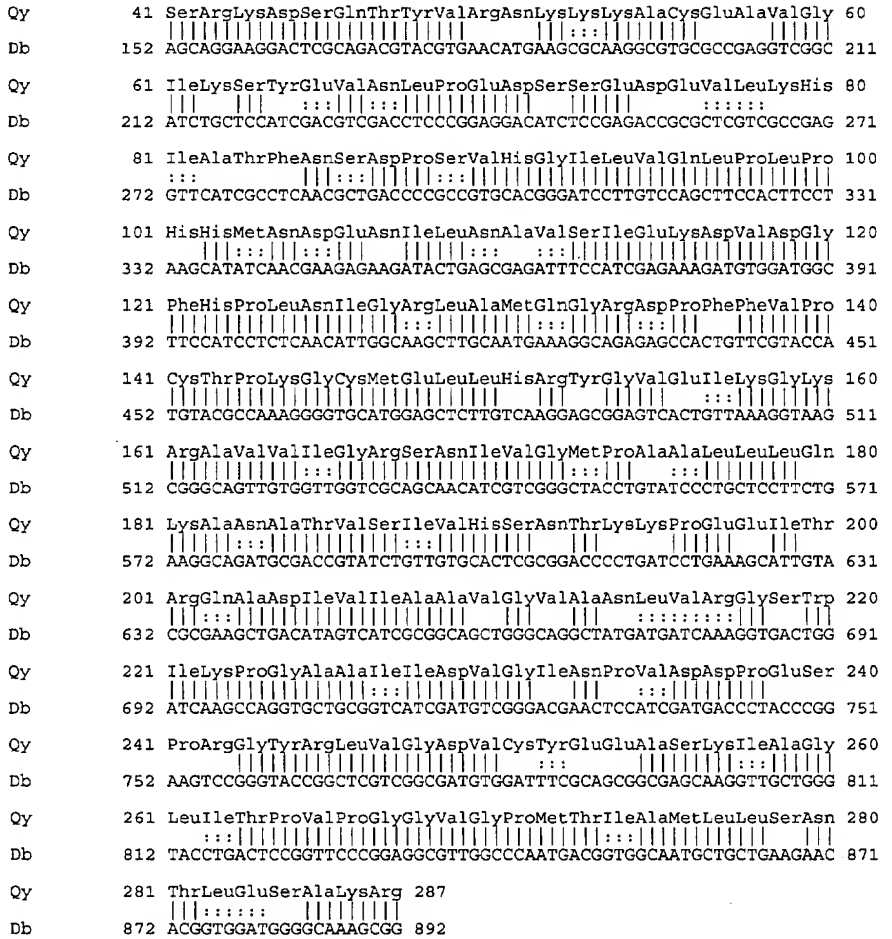
SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1143	76.9	892	14	CD438219	CD438219 EL01N0510
2	1024	68.9	1087	11	AY106093	AY106093 Zea mays
3	997	67.1	737	13	BQ123966	BQ123966 EST609542
4	992	66.8	860	13	BQ165274	BQ165274 EST611143
5	985	66.3	621	14	CD057603	CD057603 HO15K04S
6	982	66.1	832	13	BQ804363	BQ804363 WHE3553_F
c 7	982	66.1	838	14	CD574828	CD574828 UCRPT01_0
8	964	64.9	605	10	BE359332	BE359332 DG1_40_F1
9	945	63.6	633	12	BG415373	BG415373 HVSMEk000
c 10	934	62.9	870	14	CK298478	CK298478 EST761192
11	925	62.2	586	13	CA008897	CA008897 HU12H11r
12	924	62.2	595	13	BQ609202	BQ609202 BRY_5126
13	912	61.4	762	13	BQ295278	BQ295278 WHE2868_B
14	908	61.1	708	13	BQ788577	BQ788577 WHE4151_D
15	861	57.9	797	14	CB686156	CB686156 Bn01b_02o
16	854	57.5	696	10	AW776824	AW776824 EST335889
17	850	57.2	789	12	BI919739	BI919739 EST539674
18	841	56.6	648	13	BQ123426	BQ123426 EST609002
c 19	840	56.5	655	13	BU546290	BU546290 GM880009B
20	840	56.5	726	10	AW776355	AW776355 EST335420
21	836	56.3	859	14	CK294984	CK294984 EST757698
22	835	56.2	754	13	CA124209	CA124209 SCQGLR106
c 23	823	55.4	762	14	CF205415	CF205415 RR890915I
24	818	55.0	709	10	BE202855	BE202855 EST402877
25	818	55.0	902	11	AY112351	AY112351 Zea mays
26	817	55.0	716	12	BM448325	BM448325 DSA024E12
27	816	54.9	849	14	CB974765	CB974765 CAB30005_
28	815	54.8	691	14	CD813389	CD813389 BN15.019G
29	815	54.8	708	10	BE038548	BE038548 AB01A07 A
30	814	54.8	604	13	BQ994217	BQ994217 QGF6K04.y
31	811.5	54.6	877	12	BM817306	BM817306 HC108B02_
c 32	811	54.6	634	12	BJ230917	BJ230917 BJ230917
c 33	811	54.6	644	12	BJ230946	BJ230946 BJ230946
c 34	811	54.6	724	12	BJ242965	BJ242965 BJ242965
35	807	54.3	707	14	CD900176	CD900176 G174.114L
36	806	54.2	843	10	AW983356	AW983356 HVSMEg001
37	803	54.0	700	9	AL506837	AL506837 AL506837
38	801	53.9	691	12	BJ226005	BJ226005 BJ226005
39	799	53.8	717	14	CD875624	CD875624 AZO3.106H
40	798	53.7	716	13	BQ579696	BQ579696 WHE2973_E
c 41	797	53.6	739	12	BJ576118	BJ576118 BJ576118
c 42	786	52.9	754	14	CD900177	CD900177 G174.114L
43	782.5	52.7	699	14	CD813202	CD813202 BN15.001N
c 44	777	52.3	720	12	BJ250385	BJ250385 BJ250385
45	777	52.3	722	12	BJ244554	BJ244554 BJ244554

ALIGNMENTS

RESULT 1
CD438219
LOCUS CD438219 892 bp mRNA linear EST 03-JUN-2003
DEFINITION EL01N0510G03.b Endosperm_5 Zea mays cDNA, mRNA sequence.
ACCESSION CD438219
VERSION CD438219.1 GI:31353862
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 892)

```
  AUTHORS   Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
            Messing,J.
  TITLE     Sequencing of the maize endosperm ESTs
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Lai, Jinsheng
            Dr. Joachim Messing's lab
            Waksman Institute, Rutgers University
            190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
            Tel: 732-445-3801
            Fax: 732-445-5735
            Email: jlai@waksman.rutgers.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .892
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="W22"
                     /db_xref="taxon:4577"
                     /tissue_type="Endosperm of 7-23DAP"
                     /clone_lib="Endosperm_5"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

Alignment Scores:
Pred. No.:                6.81e-121      Length:         892
Score:                    1143.00        Matches:        221
Percent Similarity:       95.80%         Conservative:   7
Best Local Similarity:    92.86%         Mismatches:     10
Query Match:              76.92%         Indels:         0
DB:                       14             Gaps:           0

US-10-723-061-18 (1-292) x CD438219 (1-892)

Qy          55 AlaCysGluAlaValGlyIleLysSerTyrGluValAsnLeuProGluAspSerSerGlu 74
               ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db          10 GCGTGCGAAGCGGTCGGTATAAAGTCGTACGAGGTCCTTTTGCCTGAGGACAGCTCTGAG 69

Qy          75 AspGluValLeuLysHisIleAlaThrPheAsnSerAspProSerValHisGlyIleLeu 94
               :::||||||:::||||||||||||:::||||||||||||||||||   |||||||||
Db          70 GAGGAGGTTATCAAGCACATAGCGAGTTTCAACAGTGACCCGTCCGTGCCTGGCATCTTG 129

Qy          95 ValGlnLeuProLeuProHisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIle 114
               ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db         130 GTTCAGTTGCCCTTACCAAGGCATATGAACGATGAGAACATTCTGAATGCGGTCAGTATT 189

Qy         115 GluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArg 134
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         190 GAAAAGGATGTTGATGGCTTTCATCCGCTGAACATCGGACGGCTTGCAATGCAAGGCCGA 249

Qy         135 AspProPhePheValProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGly 154
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db         250 GATCCGTTCTTTGTACCTTGCACCCCGAAAGGATGCATGGAGTTGCTACACCGACACGGA 309

Qy         155 ValGluIleLysGlyLysArgAlaValValIleGlyArgSerAsnIleValGlyMetPro 174
               ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db         310 GTTGAAATTAAAGGGAAGAGAGCTGTTGTGATCGGACGAAGCAATATTGTTGGAACGCCC 369

Qy         175 AlaAlaLeuLeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLys 194
               ||||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         370 GCTGCATTGTTGTTGCAAAAAGCAAATGCAACTGTCAGCATTGTTCATTCACAAACCAAG 429

Qy         195 LysProGluGluIleThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsn 214
               ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db         430 AACCCTGAGCAAATAACGAGACAGGCAGATATAGTCATCGCAGCTGTTGGAGTCGCCAAC 489

Qy         215 LeuValArgGlySerTrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnPro 234
               ||||||||||||  ||||||||||||||||||  |||||||||||||||||||||||||
Db         490 TTGGTCAGAGGGGATTGGATAAAGCCTGGAGCTACTATTATTGATGTTGGCATCAATCCA 549

Qy         235 ValAspAspProGluSerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGlu 254
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         550 GTTGATGACCCAGAGAGCCCCCGCGGGTACAGGCTCGTCGGGGACGTCTGCTACGAGGAA 609

Qy         255 AlaSerLysIleAlaGlyLeuIleThrProValProGlyGlyValGlyProMetThrIle 274
               |||||||||:::||||||   ||||||||||||||||||||||||||||||||||||||
Db         610 GCCTCCAAGGTCGCCGGAGCCATCACGCCGGTCCCAGGCGGCGTCGGGCCGATGACGATC 669

Qy         275 AlaMetLeuLeuSerAsnThrLeuGluSerAlaLysArgIleHisLysPheLys 292
               ||||||||||||||||||||||||||||||||||:::|||||||||   ||||||
Db         670 GCGATGCTTCTGTCGAACACGCTCGAGTCGGCTAGACGGATCCACGGATTCAAG 723

RESULT 2
AY106093
LOCUS       AY106093                1087 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO112665 mRNA sequence.
ACCESSION   AY106093
VERSION     AY106093.1  GI:21209171
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1087)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1087)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .1087
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:636563"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

Alignment Scores:
Pred. No.:                4.93e-107      Length:         1087
Score:                    1024.00        Matches:        191
Percent Similarity:       81.53%         Conservative:   43
Best Local Similarity:    66.55%         Mismatches:     53
Query Match:              68.91%         Indels:         0
DB:                       11             Gaps:           0

US-10-723-061-18 (1-292) x AY106093 (1-1087)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||:::|||||||||||||||   :::|||      :::|||   |||:::|||   :::
Db         32 GCGCAGATCATCGATGGCAAGGCCATCGCCGCCGACGTCCGCCGCGAGGTCGCCGCCGAT 91

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::|||   :::       |||   |||:::|||||||||||||||:::::|||||
Db         92 GTGGCCGCGCTCTCGTCGGCCCACGGACTCGTGCCGGGGCTGGCCGTGGTCATCGTGGGG 151
```

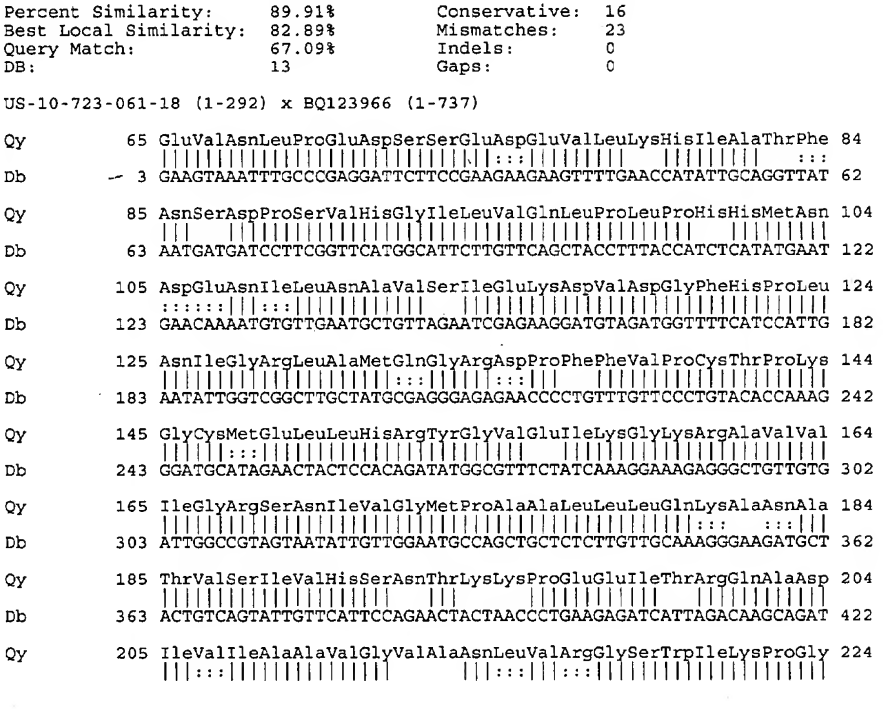


RESULT 3
BQ123966
LOCUS BQ123966 737 bp mRNA linear EST 17-APR-2002
DEFINITION EST609542 GLSD Medicago truncatula cDNA clone pGLSD-33N16, mRNA sequence.
ACCESSION BQ123966
VERSION BQ123966.1 GI:20175928
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 737)
AUTHORS Grusak,M.A., Samac,D., Town,C.D., Van Aken,S., Utterback,T., Cheung,F. and Fraser,C.M.
TITLE ESTs from late stage developing seeds of Medicago truncatula
JOURNAL Unpublished (2002)
COMMENT Contact: Grusak, M.A.
USDA/ARS Children's Nutrition Research Center
Baylor College of Medicine
1100 Bates Street, Houston, TX 77030-2600, USA
Tel: 713 798 7044
Fax: 713 798 7078
Email: mgrusak@bcm.tmc.edu
TIGR sequence name: MTRAX80TK More information is available at: www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES Location/Qualifiers
source 1. .737
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pGLSD-33N16"
/tissue_type="Immature seeds"
/dev_stage="25 to 35 days after pollination"
/lab_host="XLOLR"
/clone_lib="GLSD"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Immature seeds, collected from pods ranging in age from 25 to 35 days after pollination, were harvested from greenhouse-grown plants. Seed were removed and separated from pod walls and immediately frozen in liquid nitrogen. Seeds throughout the age range were pooled for mRNA extraction. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."
ORIGIN

Alignment Scores:

Pred. No.:	3.61e-104	Length:	737
Score:	997.00	Matches:	189
Percent Similarity:	89.91%	Conservative:	16
Best Local Similarity:	82.89%	Mismatches:	23
Query Match:	67.09%	Indels:	0
DB:	13	Gaps:	0

US-10-723-061-18 (1-292) x BQ123966 (1-737)



```
Db        423 ATCATCATCGCTGCTGTTGGCCAACCAAACATGGTTAAGGGAAGCTGGATAAAACCCGGT 482

Qy        225 AlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerProArgGlyTyr 244

Db        483 GCAGTCATTATTGATGTCGGAATCAACCCGGTAGATGATCCAAGTAGTTCTCGAGGTTAT 542

Qy        245 ArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIleThrPro 264

Db        543 AAATTGGTTGGAGATGTTTGTTATGAAGAAGCCGTAAAAGTTGCCTCTGCTATTACACCT 602

Qy        265 ValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeuGluSer 284

Db        603 GTTCCTGGAGGTGTTGGTCCAATGACTATTGCAATGCTTCTACAAAATACGCTCATCTCT 662

Qy        285 AlaLysArgIleHisLysPheLys 292

Db        663 GCAAAGAGAATCCATAGTTTTGAA 686

RESULT 4
BQ165274
LOCUS       BQ165274                 860 bp    mRNA    linear   EST 25-APR-2002
DEFINITION  EST611143 KVKC Medicago truncatula cDNA clone pKVKC-7D11, mRNA
            sequence.
ACCESSION   BQ165274
VERSION     BQ165274.1  GI:20307523
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 860)
  AUTHORS   VandenBosch,K., Endre,G., Silverstein,K, Town,C.D., Van Aken,S.,
            Utterback,T., Cheung,F. and Fraser,C.M.
  TITLE     The Medicago truncatula 'kiloclone' set; ESTs selected and
            re-arrayed from various libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
            Tel: 612 624 2755
            Fax: 612 625 1738
            Email: kvandenb@cbs.umn.edu
            TIGR sequence name: MTNAG47TK Alias Clone name:DSIR-31A20 More
            information is available at: www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .860
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="pKVKC-7D11"
                     /tissue_type="mixed tissues"
                     /dev_stage="various stages"
                     /lab_host="XLOLR"
                     /clone_lib="KVKC"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
ORIGIN

Alignment Scores:
Pred. No.:              1.72e-103      Length:         860
Score:                  992.00         Matches:        182
Percent Similarity:     83.21%         Conservative:   46
Best Local Similarity:  66.42%         Mismatches:     46

Query Match:            66.76%         Indels:         0
DB:                     13             Gaps:           0

US-10-723-061-18 (1-292) x BQ165274 (1-860)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20

Db         39 GCCACCGTAATCGACGGCAAAGCCGTGGCACAAACAATCCGATCTGAAATCACCGACGAG 98

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40

Db         99 GTCCGTCTCCTCTCTGAGAAATACGGCAAGGTTCCAGGACTTGCAGTGGTGATAGTAGGG 158

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60

Db        159 AACAGACAAGATTCAGAAAGCTATGTGAGAACAAAGAGAAAAGCTTGCGCTGATTTGGGA 218

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80

Db        219 ATAAAAAACTTCGACATTGACCTTCCTGAACAAATTTCCGAAGCTGAAGTTATCAAGCAT 278

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100

Db        279 GTTCATCAATTGAACGCTGACCCCAATGTACACGGTATATTGGTTCAACTTCCATTGCCT 338

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120

Db        339 AAGCATATAAATGAAGAGAAAGTTTTGACTGAAATCAGCCTCGAGAAGGATGTAGATGGT 398

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140

Db        399 TTTCATCCTTTGAACATCGGCAAGCTTGCCATGAAAGGCAGAGACCCTCTCTTTGTTCCC 458

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160

Db        459 TGCACTCCCAAAGCATGTCTTGAACTATTGTCACGAAGTGGTGTAAGTATAAAGGGGAAA 518

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180

Db        519 AGGGCAGTTGTGGTTGGTAGAAGCAACATAGTCGGGTTACCAGTTTCATTACTGCTTTTG 578

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200

Db        579 AAAGCAGATGCTACAGTTACCATTGCTCATTCACACACAAGTCAACCAGAAGATATCATA 638

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220

Db        639 CGTGAAGCAGATATTGTTATTGCAGCAGGACAGGCAAAGATGATCAAGGGAAGTTGG 698

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240

Db        699 ATAAAACCTGGAGCTGCAGTAATAGATGTTGGCACAAATTCTGTTGATGACCCAACTAGG 758

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260

Db        759 AAGGCAGGTTATAGACTTGTTGGAGATGTAGATTTTGAGGAAGCATCTAAAGTTGCTGGT 818

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIle 274

Db        819 TGGATTACTCCTGTTCCTGGAGGTGTAGGTCCAATGACAGTC 860

RESULT 5
CD057603
LOCUS       CD057603                 621 bp    mRNA    linear   EST 05-JUN-2003
DEFINITION  HO15K04S HO Hordeum vulgare cDNA clone HO15K04 5-PRIME, mRNA
            sequence.
ACCESSION   CD057603
VERSION     CD057603.1  GI:30601577
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
```

```
          Pooideae; Triticeae; Hordeum.
REFERENCE 1  (bases 1 to 621)
  AUTHORS Zierold,U. and Schweizer,P.
  TITLE   Barley ESTs from pathogen-attacked leaf epidermis
  JOURNAL Unpublished (2003)
COMMENT   Contact: Patrick Schweizer
          Transcriptome Analysis, Cytogenetics Department
          Institute of Plant Genetics and Crop Plant Research (IPK)
          Corrensstr. 3, D-06466 Gatersleben, Germany
          Tel: 0049 (0)39482-5660
          Fax: 0049 (0)39482-5595
          Email: schweiz@ipk-gatersleben.de
          Insert Length: 621   Std Error: 0.00
          Plate: 15  row: K  column: 4
          Seq primer: SK.
FEATURES            Location/Qualifiers
     source         1. .621
                    /organism="Hordeum vulgare"
                    /mol_type="mRNA"
                    /cultivar="Ingrid BC mlo-5"
                    /db_xref="GABI:708566"
                    /db_xref="taxon:4513"
                    /clone="HO15K04"
                    /tissue_type="leaf epidermis, 6 h and 24 h post
                    inoculation with Blumeria graminis"
                    /dev_stage="7 d after germination"
                    /lab_host="XL10-Gold"
                    /clone_lib="HO"
                    /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                    cDNA); Site_2: XhoI (3'-end of cDNA); Approximately 5 % of
                    the clones correspond to cDNA from the fungi B. graminis
                    hordei and tritici, respectively. Due to a cloning
                    artefact caused by the kit, in most cases the EcoRI site
                    is NOT present, as well as the EcoRIadapter used for
                    cloning. To excise the insert, restriction sites upstream
                    EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                    due to the cloning system used Blue/white selection for
                    recombinats is not 100% reliable. Average insert size is
                    1.2 kb"
ORIGIN

Alignment Scores:
Pred. No.:              6.79e-103       Length:         621
Score:                  985.00          Matches:        190
Percent Similarity:     95.57%          Conservative:   4
Best Local Similarity:  93.60%          Mismatches:     9
Query Match:            66.29%          Indels:         0
DB:                     14              Gaps:           0

US-10-723-061-18 (1-292) x CD057603 (1-621)

Qy         46 GlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIleLysSerTyrGlu 65
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 CAGACGTACGTGCGGAACAAGAAGAAGGCGTGCGAGGCGGTCGGTATCAAGTCGTACGAG 71

Qy         66 ValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAlaThrPheAsn 85
              ||||||||||||||||||:::||||||   ||||||:::||||||||||||||||||||
Db         72 GTCAATTTGCCGGAGGACGCTTCCGAGTACGAGGTCATCAAGCACATTGCGACCTTCAAT 131

Qy         86 SerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHisMetAsnAsp 105
              ||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        132 GATGACCCGTCTGTCCATGGCATCTTGGTTCAGTTGCCCTTGCCTCGCCATATGAACGAC 191

Qy        106 GluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHisProLeuAsn 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        192 GAGAACATATTGAATGCTGTTAGTATTGAAAAGGACGTCGATGGCTTTCATCCAGTGAAT 251

Qy        126 IleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThrProLysGly 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ATTGGACGACTGGCGATGCAAGGTCGGGATCCATTCTTTGTTCCATGCACCCCTAAAGGA 311

Qy        146 CysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAlaValValIle 165
              ||||||||||||||||||||   ||||||||||||   ||||||||||||||||||||

Db        312 TGCATGGAGCTGCTACACAGATCTGGAGTTGAAATAAATGGGAAGAGAGCTGTTGTAATT 371

Qy        166 GlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLysAlaAsnAlaThr 185
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        372 GGAAGAAGCAATATTGTGGGGACGCCTGCTGCTTTACTCTTACAAAAAGCAAATGCAACT 431

Qy        186 ValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGlnAlaAspIle 205
              ||||||||||||||||||   ||||||   ||||||||||||||||||||||||||||||
Db        432 GTGAGCATTGTACATTCAAAAACCAAGAACCCCGAGGAAATAACAAGACAAGCAGATATT 491

Qy        206 ValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLysProGlyAla 225
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 ATCATTGCGGCTGTTGGAGTTGCTAACCTGGTCAGAGGGAGTTGGATAAAGCCTGGAGCT 551

Qy        226 AlaIleIleAspValGlyIleAsnProValAspAspProGluSerProArgGlyTyrArg 245
              ||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        552 GCTATTATTGATGTTGGCATCAATCCGGTTGATGATCCAGCAAGCCCTCGAGGTTATCGT 611

Qy        246 LeuValGly 248
              |||||||||
Db        612 CTAGTTGGA 620

RESULT 6
BQ804363
LOCUS       BQ804363                 832 bp    mRNA    linear   EST 31-JUL-2002
DEFINITION  WHE3553_F06_L11ZS Wheat developing grains cDNA library Triticum
            aestivum cDNA clone WHE3553_F06_L11, mRNA sequence.
ACCESSION   BQ804363
VERSION     BQ804363.1  GI:22028492
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 832)
  AUTHORS   Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J.,
            Cronin,K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J.,
            Rausch,C.J., Wilson,C. and Woo,J.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Developing grains cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES            Location/Qualifiers
     source         1. .832
                    /organism="Triticum aestivum"
                    /mol_type="mRNA"
                    /cultivar="Butte 86"
                    /db_xref="taxon:4565"
                    /clone="WHE3553_F06_L11"
                    /tissue_type="whole grains"
                    /dev_stage="3-44 days post anthesis seed"
                    /lab_host="E. coli SOLR"
                    /clone_lib="Wheat developing grains cDNA library"
                    /note="Vector: Lambda ZAP II, excised phagemid; Site_1:
                    EcoRI; Plants were grown under six following different
                    environmental regimes in greenhouse, Environment 1)
                    24oC/17oC day/night, well-watered, with post-anthesis
                    fertilizer, Environment 2) 24oC/17oC day/night,
                    well-watered, without post-anthesis fertilizer,
                    Environment 3) 37oC/17oC day/night, well-watered, with
                    post-anthesis fertilizer, Environment 4) 37oC/17oC
                    day/night, well-watered, without post-anthesis fertilizer,
                    Environment 5) 37oC/17oC day/night plus drought, with
```

post-anthesis fertilizer, Environment 6) 37oC/17oC day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the TJ Close lab (Chin, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."

ORIGIN

Alignment Scores:

Pred. No.:	2.33e-102	Length:	832
Score:	982.00	Matches:	186
Percent Similarity:	82.90%	Conservative:	37
Best Local Similarity:	69.14%	Mismatches:	45
Query Match:	66.08%	Indels:	1
DB:	13	Gaps:	0

US-10-723-061-18 (1-292) x BQ804363 (1-832)

```
Qy         20 GluIleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuVal 39
              |||:::|||   :::      |||      :::||||||||||||||||||:::::::|||
Db          5 GAGGTCGCCGTGCTCTCGTCCGCCCACAACATCGTGCCGGGGCTGGCGGTGGTGATCGTG 64

Qy         40 GlySerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaVal 59
              |||||||||||||||||||||||||||:::   |||:::||||||||      |||
Db         65 GGGAGCAGGAAGGACTCGCAGACGTACGTGCAGATGAAGCGCAAGGCCTGCGCCGAGGTC 124

Qy         60 GlyIleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLys 79
              ||||||:::|||:::::::|||:::|||||||||||||   ||||||      ::::::
Db        125 GGCATCCGCTCCTTCGACGTCGACCTCCCCGAGGACATCTCCGAGGCCGCGCTCGTCGCC 184

Qy         80 HisIleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeu 99
                 :::      |||:::|||||:::||||||||||||||||||||||||||||||||||
Db        185 GAGGTCCACCGCCTCAACGCCGACCCCGCCGTCCACGGAATTCTTGTTCAGCTTCCATTG 244

Qy        100 ProHisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAsp 119
              |||   |||:::|||:::|||||||||||||||   :::|||||||||||||||||||||
Db        245 CCCAAGCATATCAACGAAGAAAATATCTTAAACCAGATCTCCATTGAGAAAGATGTCGAC 304

Qy        120 GlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheVal 139
              |||||||||||||||||||||:::|||||||||:::||||||||||||   ||||||
Db        305 GGCTTTCATCCTTTGAACATTGGCAAGCTCGCAATGAAAGGCAGAGATCCACTGTTCGTA 364

Qy        140 ProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGly 159
              |||||||||||||||||||||||||||||||   |||   ||||||   :::||||||
Db        365 CCTTGCACGCCAAAGGGATGCATGGAGCTCCTGTCACGAAGTGGCGTCACTGTAAAAGGA 424

Qy        160 LysArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeu 179
              |||   |||||||||:::|||||||||||||||||||:::|||   :::||||||||||
Db        425 AAACACGCAGTTGTGGTTGGGCGTAGCAACATCGTGGGTTTACCAGTATCCCTTCTCCTG 484

Qy        180 GlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIle 199
              ||||||:::|||||||||||||||||||||||||   |||      ||||||   |||
Db        485 CTGAAAGCGGACGCGACCGTGTCGATCGTGCATTCACGGACCCCAAATCCCGAAACAATT 544

Qy        200 ThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySer 219
              |||||||||||||||||||||||||||   |||   |||   :::::::::|||
Db        545 GTCCGTCAAGCAGACATTGTCATTGCAGCAGCTGGCCAGGCCATGATGATCAAGGGAGAC 604

Qy        220 TrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGlu 239
              |||||||||||||||||||:::|||||||||||||   |||   :::||||||||||
Db        605 TGGATCAAACCAGGCGCGGCGGTCATAGACGTCGGGACAAACTCCATCGACGACCCAACC 664

Qy        240 SerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAla 259
                          ||||||||||||||||||||||||   :::   ||||||||||||:::|||
Db        665 AGGAAGTCTGGGTACAGACTCGTTGGCGATGTGGATTTCTCGGAGGCGAGCAAGGTCGCG 724

Qy        260 GlyLeuIleThrProValProGlyGlyValGlyProMet-ThrIleAlaMetLeuLeuSe 279
              |||   :::|||||||||||||||||||||||||||||| |||:::||||||||||||
Db        725 GGTCACCTGACTCCGGTCCCCGGAGGCGTCGGGCCGATGAACCGTGGCGATGTTGCTGAA 784

Qy        279 rAsnThrLeuGluSerAlaLysArg 287
                 ||||||::::::   |||||||||
Db        785 GAACACGGTGGACGGCGCCAAGAGG 809
```

RESULT 7
CD574828/c

```
LOCUS       CD574828                838 bp    mRNA    linear   EST 12-JUN-2003
DEFINITION  UCRPT01_01dd09_b1 Poncirus trifoliata CTV-challenged cDNA library -
            AGI Poncirus trifoliata cDNA clone UCRPT01_01dd09, mRNA sequence.
ACCESSION   CD574828
VERSION     CD574828.1  GI:31670730
KEYWORDS    EST.
SOURCE      Poncirus trifoliata
  ORGANISM  Poncirus trifoliata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Sapindales; Rutaceae; Poncirus.
REFERENCE   1  (bases 1 to 838)
  AUTHORS   Roose,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,
            Wanamaker,S., Collura,K., Feuerbacher,O., Kim,H.R., Kudrna,D.,
            Wing,R. and Yu,Y.
  TITLE     Development of EST Resources and New Genetic Markers for California
            Citrus - Poncirus trifoliata CTV-challenged phloem - AGI
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Mikeal Roose
            Department of Botany & Plant Sciences, University of California
            Riverside, CA, 92521-0124, USA
            Tel: 9097874137
            Fax: 9097874437
            Email: mikeal.roose@ucr.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .838
                     /organism="Poncirus trifoliata"
                     /mol_type="mRNA"
                     /cultivar="Pomeroy OP"
                     /db_xref="taxon:37690"
                     /clone="UCRPT01_01dd09"
                     /tissue_type="Phloem"
                     /dev_stage="10 - 30 cm shoots"
                     /lab_host="E. coli TJC121"
                     /clone_lib="Poncirus trifoliata CTV-challenged cDNA
                     library - AGI"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse at University of California, Riverside. The
                     scion was a open-pollinated (very probably selfed)
                     seedling of Poncirus trifoliata cv Pomeroy that was
                     selected as homozygous for the Ctv resistance gene. The
                     rootstock was sweet orange infected with citrus tristeza
                     virus (CTV) isolate T514 over 1 year before sampling (CTV
                     infects sweet orange, but not genotypes carrying the Ctv
                     resistance gene. Shoots 10-30 cm long were harvested in
                     October 2000, and the green phloem (bark) was removed and
                     frozen quickly in dry ice. Total RNA was extracted using
                     TriZol reagent (Gibco). Poly(A) RNA was purified, a cDNA
                     library was made, and 0.5 million primary lambda cDNA
                     clones were in vivo excised to give a population of
                     pBluescript SK(-) phagemids. All steps to this point were
                     performed in the ML Roose lab at the University of
                     California, Riverside by X. Ye. Phagemids were plated,
                     plasmid DNA purified, cDNA clones archived, and DNA
                     sequences determined bi-directionally using an ABI3730 at
                     the Arizona Genomics Institute, University of Arizona
```

(Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wanamaker) using the HarvEST pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

ORIGIN

Alignment Scores:

Pred. No.:	2.35e-102	Length:	838
Score:	982.00	Matches:	183
Percent Similarity:	92.04%	Conservative:	25
Best Local Similarity:	80.97%	Mismatches:	18
Query Match:	66.08%	Indels:	0
DB:	14	Gaps:	0

US-10-723-061-18 (1-292) x CD574828 (1-838)

```
Qy         67 AsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAlaThrPheAsnSer 86
              :::|||||||||||||:::||||||   ||||||||||||||||||:::   ||||||
Db        837 CATTTACCTGAGGACACCTCTGAACAAGAAGTACTCAAGCATATTTCAGTCTTCAATGAT 778

Qy         87 AspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHisMetAsnAspGlu 106
              ||||||||||||||||||||||||||||||||||||||||   |||:::::::::::::
Db        777 GATCCTTCAGTTCATGGCATCCTTGTGCAGTTGCCTCTGCCTTGTCATATAGATGAGCAG 718

Qy        107 AsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHisProLeuAsnIle 126
              :::|||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        717 AGCATCTTAAATGCTGTCAGCATGGAGAAAGATGTGGACGGTTTCCACCCGCTGAATATT 658

Qy        127 GlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThrProLysGlyCys 146
              ||||||||||||||:::||||||:::|||   |||:::||||||||||||||||||||||
Db        657 GGCCGTCTAGCCATGCGAGGTAGAGAACCCTTGTTCATTCCTTGTACGCCTAAAGGTTGC 598

Qy        147 MetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAlaValValIleGly 166
              :::|||||||||||||||||||||   :::||||||||||||||||||||||||||||||
Db        597 ATAGAGTTGTTGCATAGATATGGCTTTGATATCAAAGGAAAAAGGGCTGTCGTGATTGGC 538

Qy        167 ArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLysAlaAsnAlaThrVal 186
              ||||||||||||||||||||||||||||||||||||||||:::   :::|||||||||
Db        537 CGAAGCAATATTGTTGGGATGCCAGCTGCTTTGTTGCTCCAAAGGGAAGATGCTACTGTT 478

Qy        187 SerIleValHisSerAsnThrLysLysProGluGluIleThrArgGlnAlaAspIleVal 206
              ||||||||||||||   ||||||   ||||||||||||||||||||||||||||:::
Db        477 AGTATTGTCCATTCCAGAACTAAGAACCCTGAGGAGATTACGAGACAAGCCGATATCATA 418

Qy        207 IleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLysProGlyAlaAla 226
              |||:::|||||||||      |||:::||||||||||||||||||||||||||||||||
Db        417 ATTTCCGCTGTAGGGCAACCAAATATGGTAAGGGGGAGCTGGATAAAACCTGGTGCAGTC 358

Qy        227 IleIleAspValGlyIleAsnProValAspAspProGluSerProArgGlyTyrArgLeu 246
              ||||||||||||||||||||||||||:::|||   :::||||||||||||||||||||||
Db        357 ATAATTGATGTTGGAATAAACCCAGTCGAGGATGCAAAAAGTCCCCGAGGTTACCGGTTG 298

Qy        247 ValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIleThrProValPro 266
              ||||||||||||||||||||||||||   ::::::|||      ||||||||||||||||
Db        297 GTTGGAGATGTTTGCTACGAAGAGGCATGTGAGGTTGCTTCAGCTATCACTCCAGTTCCT 238

Qy        267 GlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeuGluSerAlaLys 286
              ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db        237 GGAGGAGTTGGTCCAATGACTATAGCAATGCTGCTCTCTAATACTCTCACTTCGGCAAAG 178

Qy        287 ArgIleHisLysPheLys 292
              |||:::|||   |||:::
Db        177 AGGGTGCACAACTTCCAG 160
```

RESULT 8
BE359332
LOCUS BE359332 605 bp mRNA linear EST 20-JUL-2000
DEFINITION DG1_40_F10.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BE359332
VERSION BE359332.1 GI:9300889
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 605)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 599
POLYA=No.
FEATURES Location/Qualifiers
source 1. .605
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DG1)"
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
ORIGIN

Alignment Scores:

Pred. No.:	1.73e-100	Length:	605
Score:	964.00	Matches:	186
Percent Similarity:	98.43%	Conservative:	2
Best Local Similarity:	97.38%	Mismatches:	3
Query Match:	64.87%	Indels:	0
DB:	10	Gaps:	0

US-10-723-061-18 (1-292) x BE359332 (1-605)

```
Qy        102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 CATATGAACGATGAGAATATTTTGAATGCTGTCAGTATTGAAAAGGATGTTGATGGCTTT 68

Qy        122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 CATCCACTGAACATTGGACGACTCGCAATGCAAGGTCGAGATCCATTCTTTGTACCTTGC 128

Qy        142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 ACCCCAAAAGGATGCATGGAGTTGCTACACCGATATGGAGTTGAAATTAAAGGGAAGAGA 188

Qy        162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLys 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 GCTGTTGTAATCGGAAGAAGCAATATTGTTGGAATGCCTGCTGCATTGTTATTGCAAAAA 248

Qy        182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
              ||||||||||||||||||||||||||||||   ||||||   ||||||||||||||||||
Db        249 GCAAATGCAACTGTCAGCATTGTACATTCACAAACTAAGAACCCTGAGGAAATAACACGA 308

Qy        202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        309 CAGGCAGATATAGTCATCGCAGCAGTTGGAGTTGCCAACTTGGTCAGAGGGAATTGGATA 368

Qy        222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerPro 241
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          369 AAACCTGGAGCAGCTATTATTGATGTTGGTATCAACCCAGTTGATGACCCAGAAAGCCCT 428

Qy          242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
                ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          429 CGGGGTTATAGGCTCGTCGGAGATGTTTGTTATGAGGAAGCCTCCAAGGTTGCCGGAGCA 488

Qy          262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          489 ATCACACCAGTTCCAGGAGGTGTTGGGCCGATGACAATTGCAATGCTTTTGTCAAACACA 548

Qy          282 LeuGluSerAlaLysArgIleHisLysPheLys 292
                ||||||||||||||||||||||||||||||||
Db          549 CTTGAGTCAGCTAAAAGAATCCACAAATTCAAA 581

RESULT 9
BG415373
LOCUS       BG415373                633 bp    mRNA    linear   EST 23-OCT-2001
DEFINITION  HVSMEk0006D16f Hordeum vulgare testa/pericarp EST library
            HVcDNA0013 (normal) Hordeum vulgare subsp. vulgare cDNA clone
            HVSMEk0006D16f, mRNA sequence.
ACCESSION   BG415373
VERSION     BG415373.1  GI:13320924
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Kannangara,G., von
            Wetstein,D., Akhunov,E., Chin,A., Choi,D.W., Fenton,R.D.,
            Kianian,P., Otto,C., Simons,K., Zhang,D., Begum,D., Frisch,D.,
            Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and
            Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex testa/pericarp cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 267
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 8
            High quality sequence stop: 598.
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HVSMEk0006D16f"
                     /tissue_type="testa/pericarp"
                     /lab_host="TJC121"
                     /clone_lib="Hordeum vulgare testa/pericarp EST library
                     HVcDNA0013 (normal)"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Plants were raised from seeds in a Controlled Environments
                     growth chamber maintained in continuous light at 18oC, and
                     testa and pericarp were dissected from developing kernels
                     at Washington State University, Pullman, WA (Kannangara,
                     von Wetstein). Total RNA was prepared, poly(A) RNA was
                     purified, one cDNA library was made, and 1 million pfu
                     were in vivo excised to give pBluescript SK(-) cDNA
                     phagemids in the TJ Close lab at the University of
                     California, Riverside (Akhunov, Chin, Choi, Close, Fenton,
                     Kianian, Otto, Simons, Zhang). Phagemids were plated and
                     picked at the Clemson University Genomics Institute (CUGI)
                     (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
                     preparations, DNA sequencing and sequence analysis were
                     performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
                     Oates, Rambo, Main). The sequence has been trimmed to
                     remove vector sequence and contains a minimum of 100 bases
                     of phred value 20 or above. For more details on library
                     preparation and sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
ORIGIN

Alignment Scores:
Pred. No.:                  2.89e-98        Length:          633
Score:                      945.00          Matches:         186
Percent Similarity:         91.39%          Conservative:    5
Best Local Similarity:      89.00%          Mismatches:      18
Query Match:                63.59%          Indels:          0
DB:                         12              Gaps:            0

US-10-723-061-18 (1-292) x BG415373 (1-633)

Qy           30 ValValProGlyLeuAlaValIleLeuValGlySerArgLysAspSerGlnThrTyrVal 49
                ||||||||   ||||||   ||||||||||||||||||   |||:::   |||||||||
Db            6 GTGGTGCCTGAGCTGGCGGCCATCCTAGTCGGCTCGAGGAACGATGCCCATACTTACGTG 65

Qy           50 ArgAsnLysLysLysAlaCysGluAlaValGlyIleLysSerTyrGluValAsnLeuPro 69
                ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db           66 CGGAACAAGAAGAAGGCGGGCGAGGCGGTCGGTATCAAGTCGTACGAGGTCAATTTGCCG 125

Qy           70 GluAspSerSerGluAspGluValLeuLysHisIleAlaThrPheAsnSerAspProSer 89
                ||||||:::||||||   |||:::||||||||||||||||||   |||||||||
Db          126 GAGGACGCTTCCGAGTACTAGGTCATCAAGCACATTGCGACCTTCAATGATGACCCGTCT 185

Qy           90 ValHisGlyIleLeuValGlnLeuProLeuProHisHisMetAsnAspGluAsnIleLeu 109
                ||||||||||||   |||   ||||||||||||   ||||||||||||||||||||||
Db          186 GTCCATGGCATCTGGGTTCATTTGCCCTTGCCTCGCCATATGAACGACGAGAACATATTG 245

Qy          110 AsnAlaValSerIleGluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeu 129
                ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db          246 AATGCTGTTAGTATTGAAAAGGACGTCGATGGCTTTCATCCAGTGAATATTGGACGACTG 305

Qy          130 AlaMetGlnGlyArgAspProPhePheValProCysThrProLysGlyCysMetGluLeu 149
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          306 GCGATGCAAGGTCGGGATCCATTCTTTGTTCCATGCACCCCTAAAGGATGCATGGAGCTG 365

Qy          150 LeuHisArgTyrGlyValGluIleLysGlyLysArgAlaValValIleGlyArgSerAsn 169
                ||||||||   ||||||||||   ||||||||||||||||||||||||||||||||||||
Db          366 CTACACAGATCTGGAGTTGAAATAAATGGGAAGAGAGCTGTTGTAATTGGAAGAAGCAAT 425

Qy          170 IleValGlyMetProAlaAlaLeuLeuGlnLysAlaAsnAlaThrValSerIleVal 189
                ||||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db          426 ATTGTGGGGACGCCTGCTGCTTTACTCTTACAAAAAGCAAATGCAACTGTGAGCATTGTA 485

Qy          190 HisSerAsnThrLysLysProGluGluIleThrArgGlnAlaAspIleValIleAlaAla 209
                ||||||   ||||||   ||||||||||||||||||||||||   :::||||||||||
Db          486 CATTCAAAAACCAAAAACCCCGAGGAAATAACAAGACAAGCAGATTTTATCATTGCGGCT 545

Qy          210 ValGlyValAlaAsnLeuValArgGlySerTrpIleLysProGlyAlaAlaIleIleAsp 229
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          546 GTTGGAGTTGCTAACCTGGTCAGAGGGAGTTGGATAAAGCCTGGAGCTGCTATTATCGAT 605

Qy          230 ValGlyIleAsnProValAspAspPro 238
                ||||||||||||||||||||||||
Db          606 GGTGGCATCAATCCGGTTGATGATCCA 632

RESULT 10
CK298478/c
LOCUS       CK298478                870 bp    mRNA    linear   EST 15-DEC-2003
DEFINITION  EST761192 Nicotiana benthamiana mixed tissue cDNA library,
```

```
            normalized, full-length Nicotiana benthamiana cDNA clone NBMDG17 3'
            end, mRNA sequence.
ACCESSION   CK298478
VERSION     CK298478.1  GI:39885892
KEYWORDS    EST.
SOURCE      Nicotiana benthamiana
  ORGANISM  Nicotiana benthamiana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Nicotiana.
REFERENCE   1  (bases 1 to 870)
  AUTHORS   Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A., Day,B.,
            Staskawicz,B., Jin,H. and Baker,B.
  TITLE     Generation of EST sequences from Nicotiana benthamiana
  JOURNAL   Unpublished (2003)
COMMENT     Other_ESTs: EST761191
            Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            Clones can be requested from TIGR via potato@tigr.org
            Seq primer: GTA ATA CGA CTC ACT ATA GGG C.
FEATURES             Location/Qualifiers
     source          1. .870
                     /organism="Nicotiana benthamiana"
                     /mol_type="mRNA"
                     /db_xref="taxon:4100"
                     /clone="NBMDG17"
                     /tissue_type="abiotic and biotic stress-treated leaves,
                     callus tissue and root tissue"
                     /lab_host="DH10B-TonA"
                     /clone_lib="Nicotiana benthamiana mixed tissue cDNA
                     library, normalized, full-length"
                     /note="Vector: pCMVSport6.1; Site_1: EcoRI; Site_2: NotI;
                     supplier: RNA was isolated from Nicotiana benthamiana
                     tissues that include callus, roots from liquid culture
                     grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
                     cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
                     challenged leaves (Pseudomonas syringae pv tomato 12 hr;
                     Xanthomonas campestris pv campestris 12 hr, 18hr;
                     Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
                     campestris pv vesicatoria 18hr). RNA was isolated from
                     these tissues and pooled in approximately equal molar
                     amounts."
ORIGIN

Alignment Scores:
Pred. No.:                8.6e-97        Length:          870
Score:                    934.00         Matches:         178
Percent Similarity:       90.65%         Conservative:    16
Best Local Similarity:    83.18%         Mismatches:      20
Query Match:              62.85%         Indels:          0
DB:                       14             Gaps:            0

US-10-723-061-18 (1-292) x CK298478 (1-870)

Qy          78 LeuLysHisIleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeu 97
               ||||||:::|||:::   ||||||   |||      ||||||||||||||||||||||||
Db         870 CTCAAGTATATCTCAGGCTTCAATGATGATCTTATGGTTCATGGAATTCTTGTTCAGCTA 811

Qy          98 ProLeuProHisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAsp 117
               |||||||||   |||||||||::::::|||||||||||||||||   ||||||||||||
Db         810 CCGTTACCTTCTCATATGAATGAACAGAACATCTTAAATGCTGTTTGCATTGAGAAGGAT 751

Qy         118 ValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhe 137
               ||||||||||||||||||||||||||||||||||||||||:::||||||:::|||
Db         750 GTGGATGGATTCCACCCACTAAATATTGGCCGGCTAGCGATGCGAGGTAGAGAACCACTA 691

Qy         138 PheValProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIle 157
               ||||||||||||||||||||||||||:::||||||||||||||||||   |||||||||
Db         690 TTTGTTCCATGTACACCTAAAGGATGCATTGAGCTACTGCATCGGTATAATGTGGAGATC 631

Qy         158 LysGlyLysArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeu 177
               |||||||||||||||||||:::|||||||||||||||||||||   ||||||||||||
Db         630 AAGGGGAAAAGAGCTGTTGTTGTTGGCAGGAGTAATATCGTCGGAACGCCGGCTGCTCTG 571

Qy         178 LeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGlu 197
               |||||||||:::   :::|||||||||||||||||||||||||   ||||||||||||||
Db         570 CTTCTGCAGAGGGAAGATGCCACTGTCAGCATTGTCCATTCCAGAACCAAGAAACCTGAG 511

Qy         198 GluIleThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArg 217
               |||||||||||||||||||||||:::|||:::|||||||||     |||:::||||||
Db         510 GAGATCACGAGACAAGCAGATATCATAATCTCAGCTGTAGGTCAACCAAACATGGTTAGA 451

Qy         218 GlySerTrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAsp 237
               ||||||||||||||||||||||||   |||||||||||||||||||||||||||:::|||
Db         450 GGCAGCTGGATCAAGCCTGGGGCTGTGATTATTGATGTCGGGATCAACCCTGTTGAGGAT 391

Qy         238 ProGluSerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLys 257
                  :::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         390 GCTACAAATCCTCGCGGCTATCGGCTAGTTGGAGATGTTTGTTATGAAGAGGCCAGCAAA 331

Qy         258 IleAlaGlyLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeu 277
               :::|||     ||||||||||||||||||||||||||||||||||||||||||||||||
Db         330 GTTGCCTCAGCTATCACTCCTGTTCCTGGGGGAGTTGGACCCATGACTATAGCAATGCTT 271

Qy         278 LeuSerAsnThrLeuGluSerAlaLysArgIleHisLysPhe 291
               ||||||||||||||   ||||||||||||||||||   |||
Db         270 CTTTCTAATACTTTGTTATCAGCAAAGCGAATTCACAACTTC 229

RESULT 11
CA008897
LOCUS       CA008897                586 bp    mRNA    linear   EST 23-OCT-2002
DEFINITION  HU12H11r HU Hordeum vulgare subsp. vulgare cDNA clone HU12H11
            5-PRIME, mRNA sequence.
ACCESSION   CA008897
VERSION     CA008897.1  GI:24285879
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and
            Graner,A.
  TITLE     Barley ESTs from germinating seeds
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 586   Std Error: 0.00
            Plate: 12  row: H  column: 11
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="barke"
                     /sub_species="vulgare"
                     /db_xref="GABI:266547"
                     /db_xref="taxon:112509"
                     /clone="HU12H11"
                     /tissue_type="germinating seeds"
                     /dev_stage="germinating seeds, 16-48 h"
                     /lab_host="XL10-Gold"
                     /clone_lib="HU"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); germinating seeds,
                     16-48h. Due to a cloning artefact caused by the kit, in
                     most cases the EcoRI site is NOT present, as well as the
                     EcoRIadapter used for cloning. To excise the insert,
```

restriction sites upstream EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinats is not 100% reliable. Average insert size is 1.2 kb"

ORIGIN

Alignment Scores:

Pred. No.:	5.23e-96	Length:	586
Score:	925.00	Matches:	179
Percent Similarity:	94.87%	Conservative:	6
Best Local Similarity:	91.79%	Mismatches:	10
Query Match:	62.25%	Indels:	0
DB:	13	Gaps:	0

US-10-723-061-18 (1-292) x CA008897 (1-586)

```
Qy         42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
Db          1 CGGCACGAGGCCCAGACGTACGTGCGGAACAAGAAGAAGGCGTGCGAGGCGGTCGGTATC 60

Qy         62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
Db         61 AAGTCGTACGAGGTCAATTTGCCGGAGGACGCTTCCGAGTACGAGGTCATCAAGCACATT 120

Qy         82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
Db        121 GCGACCTTCAATGATGACCCGTCTGTCCATGGCATCTTGGTTCAGTTGCCCTTGCCTCGC 180

Qy        102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
Db        181 CATATGAACGACGAGAACATATTGAATGCTGTTAGTATTGAAAAGGACGTCGATGGCTTT 240

Qy        122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
Db        241 CATCCAGTGAATATTGGACGACTGGCGATGCAAGGTCGGGATCCATTCTTTGTTCCATGC 300

Qy        142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
Db        301 ACCCCTAAAGGATGCATGGAGCTGCTACACAGATCTGGAGTTGAAATAAATGGGAAGAGA 360

Qy        162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGlnLys 181
Db        361 GCTGTTGTAATTGGAAGAAGCAATATTGTGGGGACGCCTGCTGCTTTACTCTTACAAAAA 420

Qy        182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
Db        421 GCAAATGCAACTGTGAGCATTGTACATTCAAAAACCAAGAACCCCGAGGAAATAACAAGA 480

Qy        202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
Db        481 CAAGCAGATATTATCATTGCGGCTGTTGGAGTTGCTAACCTGGTCAGAGGGAGTTGGATA 540

Qy        222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAsp 236
Db        541 AAGCCTGNAGCTGCTATTATTGATGTTGGCATCAATCCGGTTGAT 585
```

RESULT 12
BQ609202

```
LOCUS       BQ609202                595 bp    mRNA    linear   EST 25-JUN-2002
DEFINITION  BRY_5126 wheat EST endosperm library Triticum aestivum cDNA 5',
            mRNA sequence.
ACCESSION   BQ609202
VERSION     BQ609202.1  GI:21558541
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Clarke,B., Lambrecht,M. and Rhee,S.Y.
  TITLE     Arabidopsis genomic information for interpreting wheat EST
            sequences
  JOURNAL   Funct. Integr. Genomics 3 (1-2), 33-38 (2003)
  MEDLINE   22478026
  PUBMED    12590341
COMMENT     Contact: Lambrecht M
            The Arabidopsis Information Resource
            Carnegie Institution of Washington, Dept. of Plant Biology
            260 Panama Street, Stanford, CA 94305, USA
            Tel: 1 650 325 1521 x 251
            Fax: 1 650 325 3748
            Email: rhee@acoma.stanford.edu.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Wyuna"
                     /db_xref="taxon:4565"
                     /tissue_type="endosperm"
                     /dev_stage="developing endosperm tissue 8, 10 and 12 DPA
                     (days post anthesis)"
                     /clone_lib="wheat EST endosperm library"
ORIGIN
```

Alignment Scores:

Pred. No.:	6.98e-96	Length:	595
Score:	924.00	Matches:	182
Percent Similarity:	93.97%	Conservative:	5
Best Local Similarity:	91.46%	Mismatches:	10
Query Match:	62.18%	Indels:	2
DB:	13	Gaps:	0

US-10-723-061-18 (1-292) x BQ609202 (1-595)

```
Qy         72 SerSerGluAspGluValLeuLysHisIleAlaThrPheAsnSerAspProSerValHis 91
Db          3 GCTTCCGAGTACGAGGTCATCAAGCACATAGCAACCTTCAACGATGACCCGTCGGTGCAT 62

Qy         92 GlyIleLeuValGlnLeuProLeuProHisHisMetAsnAspGluAsnIleLeuAsnAla 111
Db         63 GGCATCCTGGTTCAGCTGCCCTTACCTCACCATATGAACGACGAGAACATATTGAATGCT 122

Qy        112 ValSerIleGluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMet 131
Db        123 GTTAGTATTGAAAAGGATGTTGATGGCTTTCATCCAGTGAACATTGGACGACTTGCGATG 182

Qy        132 GlnGlyArgAspProPhePheValProCysThrProLysGlyCysMetGluLeuLeuHis 151
Db        183 CAAGGTCGGGATCCATTCTTTGTTCCATGCACCCCTAAAGGATGCATGGAGCTGCTACAC 242

Qy        152 ArgTyrGlyValGluIleLysGlyLysArgAlaValValIleGlyArgSerAsnIleVal 171
Db        243 AGATCTGGAGTTGAAATAAAAGGGAAGAGAGCTGTTGTAATTGGAAGGAGCAATATTGTG 302

Qy        172 GlyMetProAlaAlaLeuLeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSer 191
Db        303 GGGACGCCTGCTGCTTTACTCTTACAAAAAGCAAATGCAACTGTGAGTATTGTACATTCA 362

Qy        192 AsnThrLysLysProGluGluIleThrArgGlnAlaAspIleValIleAlaAlaValGly 211
Db        363 AAAACCAAGAACCCCGAGGAAATAACAAGACAAGCAGATATTATAATCGCGGCTGTTGGA 422

Qy        212 ValAlaAsnLeuValArgGlySerTrpIleLysProGlyAlaAlaIleIleAspValGly 231
Db        423 GTTGCTAACCTGGTCAGAGGGAGTTGGATAAAGCCTGGAGCTGCTATTATTGATGTTGGC 482

Qy        232 IleAsnProValAspAspProGluSerProArgGlyTyrArgLeuValGlyAspValCys 251
Db        483 ATCAATCCGGTTGATGATCCAGCAAGTCCTCGAGGTTACCGGCTAGTTGGAGATGTCTGC 542

Qy        252 TyrGluGluAlaSerLys-IleAlaGlyLeuIleThrPro-ValProGlyGly 268
Db        543 TTTGAGGAAGCCTNCAAGGGNTGCAGGAGCCATCACACCAGGTCCCCGGCGGG 595
```

RESULT 13

```
BQ295278
LOCUS       BQ295278                762 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  WHE2868_B10_C20ZS Wheat unstressed root tip cDNA library Triticum
            aestivum cDNA clone WHE2868_B10_C20, mRNA sequence.
ACCESSION   BQ295278
VERSION     BQ295278.1  GI:20810800
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 762)
  AUTHORS   Anderson,O.D., Chao,S., Chin,A., Close,T.J., Crossman,C.,
            Gustafson,P., Lazo,G.R., Pham,J., Rausch,C.J., Ross,K., Wilson,C.
            and Woo,J.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Unstressed root tip cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES             Location/Qualifiers
     source          1. .762
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE2868_B10_C20"
                     /tissue_type="Root tip at 1.0 to 1.5 mm stage"
                     /dev_stage="Four-day old seedling"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat unstressed root tip cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown under
                     hydroponic conditions for four days. Root tips were
                     excised and snap frozen (Ross and Gustafson) and total RNA
                     was prepared at University of Missouri, Columbia. Poly(A)
                     RNA was purified, a cDNA library was made, and the cDNA
                     clones were in vivo excised to give pBluescript
                     SK(-)phagemids in the TJ Close lab (Chin and Close) at the
                     University of California, Riverside. Plasmid DNA
                     preparations and DNA sequencing were performed in the OD
                     Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:              2.44e-94        Length:         762
Score:                  912.00          Matches:        178
Percent Similarity:     96.86%          Conservative:   7
Best Local Similarity:  93.19%          Mismatches:     6
Query Match:            61.37%          Indels:         0
DB:                     13              Gaps:           0

US-10-723-061-18 (1-292) x BQ295278 (1-762)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||:::   ||||||
Db        189 GCAAAGATCATCGACGGGAAGTTGGTGGCGAAGCAAATAAGAGAGGAGGTCGGCGTTGAG 248

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 ATCGCCCGGATGAAGGACGCGATCGGGGTGGTGCCTGGGCTGGCGGTCATCCTAGTCGGC 308

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||:::||||||||||
Db        309 TCGAGGAAGGATTCCCAGACCTACGTGCGGAACAAGAAGAGGGCCTGTGATGCGGTCGGT 368

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              ||||||||||||||||||||||||||||||||||:::||||||   ||||||:::||||||
Db        369 ATCAAGTCGTATGAGGTCAATCTGCCAGAGGACGCTTCCGAGTACGAGGTCATCAAGCAC 428

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        429 ATAGCAACCTTCAACGATGATCCGTCGGTGCATGGCATCTTGGTTCAGCTGCCCTTACCT 488

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 CGCCATATGAACGACGAGAACATATTGAATGCTGTTAGTATTGAGAAGGACGTTGATGGC 548

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        549 TTTCATCCAGTGAACATTGGACGGCTTGCGATGCAAGGTCGGGATCCATTCTTTGTTCCA 608

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db        609 TGCACCCCTAAAGGATGCATGGAGCTGCTACACAGATCTGGAGTTGAAATAAAAGGGAAG 668

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGluGln 180
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        669 AGAGCTGTTGTAATTGGAAGGAGCAATATTGTGGGGACACCTGCTGCTTTACTCTTACAA 728

Qy        181 LysAlaAsnAlaThrValSerIleValHisSer 191
              ||||||||||||||||||||||||||||||||
Db        729 AAAGCAAATGCAACTGTGAGTATTGTACATTCA 761

RESULT 14
BQ788577
LOCUS       BQ788577                708 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  WHE4151_D10_G19ZS Wheat CS whole plant cDNA library Triticum
            aestivum cDNA clone WHE4151_D10_G19, mRNA sequence.
ACCESSION   BQ788577
VERSION     BQ788577.1  GI:21997049
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 708)
  AUTHORS   Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K.,
            Dvorak,J., Lazo,G.R., Rausch,C.J., Wilson,C. and Woo,J.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Chinese Spring whole plant cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES             Location/Qualifiers
     source          1. .708
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE4151_D10_G19"
                     /tissue_type="Roots, leaves, crown, stem and sheath"
                     /dev_stage="Adult"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat CS whole plant cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid
                     pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Plant
                     tissues from wheat cv. CS grown to full tillering stage in
```

greenhouse were collected at University of California, Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total RNA was prepared from leaves (young leaf and third leaf), whole roots, crown, steam and sheath tissues, and then equal quantities of RNA were pooled from the these samples. PolyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak) at the University of California, Davis. Colony plating, plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Alignment Scores:

Pred. No.:	6.34e-94	Length:	708
Score:	908.00	Matches:	177
Percent Similarity:	96.84%	Conservative:	7
Best Local Similarity:	93.16%	Mismatches:	6
Query Match:	61.10%	Indels:	0
DB:	13	Gaps:	0

US-10-723-061-18 (1-292) x BQ788577 (1-708)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||||||||||||||||||||||||||||||||||||||||||:::   ||||||
Db        138 GCAAAGATCATCGACGGGAAGTTGGTGGCGAAGCAAATAAGAGAGGAGGTCGGCGTTGAG 197

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 ATCGCCCGGATGAAGGACGCGATCGGGGTGGTGCCTGGGCTGGCGGTCATCCTAGTCGGC 257

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||||||||||||:::||||||:::|||||||||
Db        258 TCGAGGAAGGATTCCCAGACCTACGTGCGGAACAAGAAGAGGGCCTGTGATGCGGTCGGT 317

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              ||||||||||||||||||||||||||||||:::||||||  ||||||:::||||||
Db        318 ATCAAGTCGTATGAGGTCAATCTGCCAGAGGACGCTTCCGAGTACGAGGTCATCAAGCAC 377

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        378 ATAGCAACCTTCAACGATGATCCGTCGGTGCATGGCATCTTGGTTCAGCTGCCCTTACCT 437

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 CGCCATATGAACGACGAGAACATATTGAATGCTGTTAGTATTGAGAAGGACGTTGATGGC 497

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 TTTCATCCAGTGAACATTGGACGGCTTGCGATGCAAGGTCGGGATCCATTCTTTGTTCCA 557

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        558 TGCACCCCTAAAGGATGCATGGAGCTGCTACACAGATCTGGAGTTGAAATAAAAGGGAAG 617

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGluGln 180
              ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        618 AGAGCTGTTGTAATTGGAAGGAGCAATATTGTGGGGACACCTGCTGCTTTACTCTTACAA 677

Qy        181 LysAlaAsnAlaThrValSerIleValHis 190
              |||||||||||||||||||||||||||||
Db        678 AAAGCAAATGCAACTGTGAGTATTGTACAT 707
```

RESULT 15
CB686156
LOCUS CB686156 797 bp mRNA linear EST 09-APR-2003
DEFINITION Bn01b_02o22_A
Bn01b_AAFC_ECORC_transgenic_Brassica_napus_overexpressing_BNCBF17_constitutively_frost_tolerant Brassica napus cDNA clone Bn01b_02o22, mRNA sequence.
ACCESSION CB686156
VERSION CB686156.1 GI:29689881
KEYWORDS EST.
SOURCE Brassica napus (rape)
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 797)
AUTHORS Singh,J., Allard,G., Tinker,N., Robert,L., Lacroix,C., De Moors,A., Chagnon,J., Farah,S., Couroux,P. and Hattori,J.
TITLE Expressed Sequence Tags from constitutively frost tolerant transgenic Brassica napus overexpressing BNCBF17
JOURNAL Unpublished (2002)
COMMENT Contact: Singh,J.A.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
KW Neatby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A 0C6, Canada
Tel: (613) 759-1662
Fax: (613) 759-1701
Email: singhja@em.agr.ca.
FEATURES Location/Qualifiers
source 1. .797
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Westar"
/db_xref="taxon:3708"
/clone="Bn01b_02o22"
/tissue_type="fourth leaf"
/dev_stage="3 weeks seedling grown at room temperature"
/clone_lib="Bn01b_AAFC_ECORC_transgenic_Brassica_napus_overexpressing_BNCBF17_constitutively_frost_tolerant"
/note="Vector: Bluescript SK+/XhoI-EcoRI; Site_1: EcoRI; Site_2: XhoI; Germinated in soil flats and seedlings grown for 3 weeks in a Conviron E-15 cabinet set at 20oC /16 hr light (250 Em-2sec-1) and 16 oC / 8 hr dark. Fourth leaves collected at 9 am and immediately frozen."
ORIGIN

Alignment Scores:

Pred. No.:	2e-88	Length:	797
Score:	861.00	Matches:	171
Percent Similarity:	81.67%	Conservative:	34
Best Local Similarity:	68.13%	Mismatches:	44
Query Match:	57.94%	Indels:	3
DB:	14	Gaps:	0

US-10-723-061-18 (1-292) x CB686156 (1-797)

```
Qy         40 GlySerArgLysAsp-SerGlnThrTyrValArgAsnLysLys-LysAlaCysGluAlaV 59
              |||:::|||||||||| ||||||||||||||||      |||::: |||||||||      |
Db          1 GGTGCACGGAAAGATTTCGCAGACTTATGTGAATATGAAGAGGAAAGGCGTGTGCTGAGG 60

Qy         59 alGlyIleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuL 79
              ||||||||||||||   :::|||:::|||||||   |||   ||||||   :::::::::
Db         61 TTGGGATCAAATCAGTTGACGTGGACCTACCAGRGGATGTTTCTGAAGCTGATCTTATCA 120

Qy         79 ysHisIleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProL 99
                 :::          ||||||||||||   ||||||||||||||||||||||||||||||
Db        121 GCAAAGTTCATCAACTCAATTCAGATCCTGATGTCCACGGTATATTAGTTCAACTCCCAT 180

Qy         99 euProHisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValA 119
              |||||   |||:::|||   ||||||:::|||:::|||:::||||||:::||||||||||
Db        181 TGCCGAAACATATCAATGRGGAGAATGTTTTAAGTGCAATCAGCATTGATAAAGATGTAG 240

Qy        119 spGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheV 139
              ||||||||||||||||||||||||||:::|||||||||:::||||||:::|||   |||:
Db        241 ATGGCTTCCATCCTCTGAATATTGGTAAGCTGGCCATGAAAGGCAGAGAGCCCCTTTTCC 300

Qy        139 alProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysG 159
              ::||||||||||||||||||:::|||||||||   |||   |||||||:::|||||||||
Db        301 TTCCCTGCACCCCAAAGGGATGTTTGGAACTCTTAAAGAGAAGTGGCGTAAAGATAAAAG 360

Qy        159 lyLysArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuL 179
```

```
                ||:::|||||||||||||:::|||||||||||||||||||||||:::|||   :::|||||||
Db        361 GGCAACGAGCAGTTGTAGTAGGTCGGAGTAACATCGTTGGATTACCTGTTTCGCTTCTAT 420

Qy        179 euGlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluI 199
                ||   ||||||:::||||||||||:::   ||||||||||:::||||||   ||||||   |
Db        421 TGCTCAAGGCTGATGCTACTGTTACAACTGTACATTCTCACACCAAAGATCCTGAGGCTA 480

Qy        199 leThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlyS 219
                ||   |||:::|||||||||||||||||||||   |||   |||   :::::::::
Db        481 TCATACGGGAAGCTGACATCGTTATAGCTGCAGCCGGACAAGCCTGCATGATTAAGAGCG 540

Qy        219 erTrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProG 239
                ||||||||||||||||||||:::|||||||||||||   |||   |||   ||||||
Db        541 ACTGGATTAAACCAGGGGCTGCGGTGATCGATGTTGGAACCAATGCAGTGAGTGACCCGA 600

Qy        239 luSerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleA 259
                      ||||||||||||||||||||||   |||   ||||||   |||   |
Db        601 GCAGGAAGTCAGGATACCGGTTGGTTGGAGATGTGGATTACGCAGAAGCTTGTAAAGT-G 659

Qy        259 laGlyLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuS 279
                |||||   ||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        660 CAGGTTGCATAACTCCTGTCCCTGGTGGTGTAGGTCCAATGACAGTGGCGATGCTGCTGA 719

Qy        279 erAsnThrLeuGluSerAlaLysArgIle 288
                |||||||||:::   |||||||||:::
Db        720 GGAACACCTTAGACGGTGCCAAGCGGGTC 748
```

Search completed: July 24, 2004, 09:45:06
Job time : 3415 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 24, 2004, 06:39:06 ; Search time 969 Seconds
(without alignments)
94.366 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1486	100.0	292	9	US-09-903-814-18	Sequence 18, Appl
2	1486	100.0	292	16	US-10-723-061-18	Sequence 18, Appl
3	1234	83.0	342	12	US-10-424-599-234064	Sequence 234064,
4	1030	69.3	242	9	US-09-903-814-20	Sequence 20, Appl
5	1030	69.3	242	16	US-10-723-061-20	Sequence 20, Appl
6	1025	69.0	292	12	US-10-425-114-60605	Sequence 60605, A
7	1025	69.0	299	12	US-10-425-114-56838	Sequence 56838, A
8	1019	68.6	361	12	US-10-424-599-257130	Sequence 257130,
9	1019	68.6	373	12	US-10-425-114-45924	Sequence 45924, A
10	1016	68.4	290	9	US-09-903-814-16	Sequence 16, Appl
11	1016	68.4	290	16	US-10-723-061-16	Sequence 16, Appl
12	1012	68.1	292	16	US-10-437-963-176965	Sequence 176965,
13	1010	68.0	294	12	US-10-424-599-268535	Sequence 268535,
14	994	66.9	321	12	US-10-425-114-44200	Sequence 44200, A
15	988	66.5	395	16	US-10-437-963-125501	Sequence 125501,
16	965	64.9	396	12	US-10-425-114-65365	Sequence 65365, A
17	919	61.8	605	16	US-10-437-963-155078	Sequence 155078,
18	801	53.9	220	12	US-10-425-114-72298	Sequence 72298, A
19	775	52.2	279	15	US-10-369-493-10151	Sequence 10151, A
20	735	49.5	935	16	US-10-408-765A-33	Sequence 33, Appl
21	730	49.1	344	15	US-10-369-493-16864	Sequence 16864, A
22	724	48.7	285	10	US-09-769-787-31	Sequence 31, Appl
23	722.5	48.6	297	15	US-10-369-493-21822	Sequence 21822, A
24	716	48.2	279	15	US-10-369-493-17359	Sequence 17359, A
25	715	48.1	299	15	US-10-369-493-12212	Sequence 12212, A
26	711	47.8	285	15	US-10-369-493-19483	Sequence 19483, A
27	710.5	47.8	280	15	US-10-369-493-19300	Sequence 19300, A
28	710	47.8	316	12	US-10-424-599-268512	Sequence 268512,
29	707.5	47.6	294	15	US-10-369-493-18390	Sequence 18390, A
30	701	47.2	289	9	US-09-903-814-22	Sequence 22, Appl
31	701	47.2	289	16	US-10-723-061-22	Sequence 22, Appl
32	694.5	46.7	278	15	US-10-369-493-17921	Sequence 17921, A
33	690.5	46.5	299	15	US-10-369-493-8083	Sequence 8083, Ap
34	687.5	46.3	284	15	US-10-369-493-10872	Sequence 10872, A
35	684.5	46.1	280	15	US-10-369-493-16550	Sequence 16550, A
36	683.5	46.0	280	15	US-10-369-493-7153	Sequence 7153, Ap
37	683.5	46.0	283	15	US-10-369-493-23202	Sequence 23202, A
38	681	45.8	285	15	US-10-369-493-10348	Sequence 10348, A
39	679	45.7	288	15	US-10-369-493-744	Sequence 744, App
40	678.5	45.7	279	15	US-10-369-493-4398	Sequence 4398, Ap
41	675	45.4	282	15	US-10-369-493-19921	Sequence 19921, A
42	673.5	45.3	286	15	US-10-369-493-20586	Sequence 20586, A
43	670	45.1	283	15	US-10-369-493-14053	Sequence 14053, A
44	668	45.0	284	15	US-10-369-493-18751	Sequence 18751, A
45	668	45.0	291	15	US-10-369-493-146	Sequence 146, App

ALIGNMENTS

RESULT 1
US-09-903-814-18
; Sequence 18, Application US/09903814
; Patent No. US20020102689A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB-1179-A
; CURRENT APPLICATION NUMBER: US/09/903,814
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/092,869
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-903-814-18

Query Match 100.0%; Score 1486; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.6e-145;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292


RESULT 2
US-10-723-061-18
; Sequence 18, Application US/10723061
; Publication No. US20040132150A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV1
;  CURRENT APPLICATION NUMBER: US/10/723,061
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: US 09/903,814
;  PRIOR FILING DATE: 2001-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
;   LENGTH: 292
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-10-723-061-18

  Query Match             100.0%;  Score 1486;  DB 16;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-145;
  Matches  292;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292


RESULT 3
US-10-424-599-234064
; Sequence 234064, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234064
;   LENGTH: 342
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_53384C.1.pep
US-10-424-599-234064

  Query Match             83.0%;  Score 1234;  DB 12;  Length 342;
  Best Local Similarity   79.8%;  Pred. No. 3e-119;
  Matches  233;  Conservative   29;  Mismatches   30;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              ||:|||| ||||||:||  |:::|:::|||:||||||||| |||| ||||||||||:||
Db         51 AKVIDGKSVAKQIRDEITAEVSRMRESIGVIPGLAVILVGDRKDSATYVRNKKKACESVG 110

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | | | ||||||:|:||| :|| :| ||||||||||||| |||::|||||| |||||||
Db        111 INSLEANLPEDSTEEEVLNYIAGYNDDPSVHGILVQLPLPSHMNEQNILNAVRIEKDVDG 170

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              ||||||||||:||:| |||||||||:|||||| | ||||||||||||||||||||||||
Db        171 FHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNVSIKGKRAVVIGRSNIVGMPAALLLQ 230

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              : :|||||||| |  |||| |||||:||||| ||:|||||||||| |||||||||:|| |
Db        231 REDATVSIVHSRTSNPEEIIRQADIIIAAVGQANMVRGSWIKPGAVIIDVGINPVEDPNS 290

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
              ||||:|||||||||| :||  :|||||||||||||||| ||| ||||:| |:
Db        291 PRGYKLVGDVCYEEAIRIASAVTPVPGGVGPMTIAMLLQNTLISAKRVHHFE 342


RESULT 4
US-09-903-814-20
; Sequence 20, Application US/09903814
; Patent No. US20020102689A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB-1179-A
;  CURRENT APPLICATION NUMBER: US/09/903,814
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
;   LENGTH: 242
;   TYPE: PRT
;   ORGANISM: Glycine max
US-09-903-814-20

  Query Match             69.3%;  Score 1030;  DB 9;  Length 242;
  Best Local Similarity   80.5%;  Pred. No. 2.8e-98;
  Matches  194;  Conservative   22;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         52 KKKACEAVGIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNA 111
              :|||||:||| | | ||||:|:|:||| :|| :| |||||||||||||| |||::|||||
Db          2 EKKACESVGINSLEANLPENSTEEEVLNYIAGYNDDPSVHGILVQLPLPSHMNEQNILNA 61

Qy        112 VSIEKDVDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIV 171
              | |||||||||||||||:||:| ||||||||||:|||||| | ||||||||||||||
```

```
Db         62 VRIEKDVDGFHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNVSIKGKRAVVIGRSNIV 121

Qy        172 GMPAALLLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVG 231
              |||||||||: :|||||||| |  |||| |||||:||||| ||:|||||||||| |||||
Db        122 GMPAALLLQREDATVSIVHSRTSNPEEIIRQADIIIAAVGQANMVRGSWIKPGAVIIDVG 181

Qy        232 INPVDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKF 291
              ||||:|| |||||:|||||||||| :||  :|||||||||||||| ||| ||||:| |
Db        182 INPVEDPNSPRGYKLVGDVCYEEAIRIASAVTPVPGGVGPMTIAMLLQNTLISAKRVHNF 241

Qy        292 K 292
              :
Db        242 E 242


RESULT 5
US-10-723-061-20
; Sequence 20, Application US/10723061
; Publication No. US20040132150A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV1
;  CURRENT APPLICATION NUMBER: US/10/723,061
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: US 09/903,814
;  PRIOR FILING DATE: 2001-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
;   LENGTH: 242
;   TYPE: PRT
;   ORGANISM: Glycine max
US-10-723-061-20

  Query Match            69.3%;  Score 1030;  DB 16;  Length 242;
  Best Local Similarity  80.5%;  Pred. No. 2.8e-98;
  Matches  194;  Conservative   22;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         52 KKKACEAVGIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNA 111
              :|||||:||| | | ||||:|:|:||| :|| :| |||||||||||||| |||::|||||
Db          2 EKKACESVGINSLEANLPENSTEEEVLNYIAGYNDDPSVHGILVQLPLPSHMNEQNILNA 61

Qy        112 VSIEKDVDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIV 171
              | |||||||||||||||||:||:| |||||||||:|||||| | ||||||||||||||
Db         62 VRIEKDVDGFHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNVSIKGKRAVVIGRSNIV 121

Qy        172 GMPAALLLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVG 231
              |||||||||: :|||||||| |  |||| |||||:||||| ||:|||||||||| |||||
Db        122 GMPAALLLQREDATVSIVHSRTSNPEEIIRQADIIIAAVGQANMVRGSWIKPGAVIIDVG 181

Qy        232 INPVDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKF 291
              ||||:|| |||||:|||||||||| :||  :|||||||||||||| ||| ||||:| |
Db        182 INPVEDPNSPRGYKLVGDVCYEEAIRIASAVTPVPGGVGPMTIAMLLQNTLISAKRVHNF 241

Qy        292 K 292
              :
Db        242 E 242


RESULT 6
US-10-425-114-60605
; Sequence 60605, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60605
;   LENGTH: 292
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3597-067-D5_FLI.pep
US-10-425-114-60605

  Query Match            69.0%;  Score 1025;  DB 12;  Length 292;
  Best Local Similarity  66.9%;  Pred. No. 1.2e-97;
  Matches  192;  Conservative   42;  Mismatches   53;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |:||||| :|  :| |:| ::| :  | |:||||||::||||||||||  |:|||  ||
Db          2 AQIIDGKAIAADVRREVAADVAALSSAHGLVPGLAVVIVGSRKDSQTYVNMKRKACAEVG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | | :|:|||| |||  ::  :   |:||:|||||||||| |:|:| ||: :|||||||
Db         62 ICSIDVDLPEDISETALVAEVHRLNADPAVHGILVQLPLPKHINEEKILSEISIEKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:||:| |||||||||||| | || :|||||||:|||||||:| :|||
Db        122 FHPLNIGKLAMKGREPLFVPCTPKGCMELLSRSGVTVKGKRAVVVGRSNIVGLPVSLLLL 181

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:||||:||| |  || | |:|||||| | | :::| |||||||:|||| | ||||
Db        182 KADATVSVVHSRTPDPESIVREADIVIAAAGQAMMIKGDWIKPGAAVIDVGTNSVDDPTR 241

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                |||||||  :  |||:|| :||||||||||:|||| ||:: |||
Db        242 KSGYRLVGDVDFAAASKVAGYLTPVPGGVGPMTVAMLLKNTVDGAKR 288


RESULT 7
US-10-425-114-56838
; Sequence 56838, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56838
;   LENGTH: 299
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMFLMO17055A08_FLI.pep
US-10-425-114-56838

  Query Match            69.0%;  Score 1025;  DB 12;  Length 299;
  Best Local Similarity  66.9%;  Pred. No. 1.3e-97;
  Matches  192;  Conservative   42;  Mismatches   53;  Indels    0;  Gaps    0;
```

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |:||||| :|  :| |:| ::| :  | |:||||||::||||||||||  |:|||  ||
Db          9 AQIIDGKAIAADVRREVAADVAALSSAHGLVPGLAVVIVGSRKDSQTYVNMKRKACAEVG 68

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | | :|:|||| ||  ::  :  |:||:|||||||||||| |:|:| ||: :||||||||
Db         69 ICSIDVDLPEDISETALVAEVHRLNADPAVHGILVQLPLPKHINEEKILSEISIEKDVDG 128

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:||:| |||||||||||||| | || :|||||||:|||||||:| :|||
Db        129 FHPLNIGKLAMKGREPLFVPCTPKGCMELLSRSGVTVKGKRAVVVGRSNIVGLPVSLLLL 188

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:||||:||| |  || | |:||||||| | | :::| ||||||:|||| | ||||
Db        189 KADATVSVVHSRTPDPESIVREADIVIAAAGQAMMIKGDWIKPGAAVIDVGTNSVDDPTR 248

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
              ||||||||  :  |||:|| :||||||||||||:|||| ||:: |||
Db        249 KSGYRLVGDVDFAAASKVAGYLTPVPGGVGPMTVAMLLKNTVDGAKR 295


RESULT 8
US-10-424-599-257130
; Sequence 257130, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 257130
;   LENGTH: 361
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_74210C.1.pep
US-10-424-599-257130

  Query Match             68.6%;  Score 1019;  DB 12;  Length 361;
  Best Local Similarity   65.4%;  Pred. No. 7.2e-97;
  Matches  189;  Conservative   47;  Mismatches   53;  Indels    0;  Gaps    0;

Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
              ::||||:: :|| :|| :: :||  :| |||||||||| |:|||||||  ||| ||||
Db         72 VLDGKLISMEIRSKIAAKVRQMKKGLGKVPGLAVILVGQRRDSQTYVRNKIMACEEVGIK 131

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
              |    || | :  :|   |  || |||:|||||||||| |:::| :|:|| :||||||||
Db        132 SLVTELPTDCAVTDVQNAIMRFNKDPSIHGILVQLPLPQHLDEEKVLDAVCLEKDVDGFH 191

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
              |||:| ||::||:| | |||||||:||| | |||| ||:||||| |||||:||:||||:
Db        192 PLNMGNLAIRGREPLFTPCTPKGCIELLIRSGVEIMGKKAVVIGTSNIVGLPASLLLQRH 251

Qy        183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
              :|||:::|: |: ||:|| :||||::| || |||||:|||||||:|||  ||:||
Db        252 HATVTVIHAFTQNPEQITSEADIVVSAAGVPNLVRGNWIKPGAAVIDVGTTPVEDPGCED 311

Qy        243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKF 291
              |||| |||||||| |:| :|||||||||||:|||| |||:|||:  |
Db        312 GYRLAGDVCYEEAVKVASIITPVPGGVGPMTVAMLLCNTLDSAKRMLNF 360


RESULT 9
US-10-425-114-45924
; Sequence 45924, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45924
;   LENGTH: 373
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 701050070_FLI.pep
US-10-425-114-45924

  Query Match             68.6%;  Score 1019;  DB 12;  Length 373;
  Best Local Similarity   65.4%;  Pred. No. 7.6e-97;
  Matches  189;  Conservative   47;  Mismatches   53;  Indels    0;  Gaps    0;

Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
              ::||||:: :|| :|| :: :||  :| |||||||||| |:|||||||  ||| ||||
Db         84 VLDGKLISMEIRSKIAAKVRQMKKGLGKVPGLAVILVGQRRDSQTYVRNKIMACEEVGIK 143

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
              |    || | :  :|   |  || |||:|||||||||| |:::| :|:|| :||||||||
Db        144 SLVTELPTDCAVTDVQNAIMRFNKDPSIHGILVQLPLPQHLDEEKVLDAVCLEKDVDGFH 203

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
              |||:| ||::||:| | |||||||:||| | |||| ||:||||| |||||:||:||||:
Db        204 PLNMGNLAIRGREPLFTPCTPKGCIELLIRSGVEIMGKKAVVIGTSNIVGLPASLLLQRH 263

Qy        183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
              :|||:::|: |: ||:|| :||||::| || |||||:|||||||:|||  ||:||
Db        264 HATVTVIHAFTQNPEQITSEADIVVSAAGVPNLVRGNWIKPGAAVIDVGTTPVEDPGCED 323

Qy        243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKF 291
              |||| |||||||| |:| :|||||||||||:|||| |||:|||:  |
Db        324 GYRLAGDVCYEEAVKVASIITPVPGGVGPMTVAMLLCNTLDSAKRMLNF 372


RESULT 10
US-09-903-814-16
; Sequence 16, Application US/09903814
; Patent No. US20020102689A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB-1179-A
;  CURRENT APPLICATION NUMBER: US/09/903,814
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
;   LENGTH: 290
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-903-814-16

  Query Match             68.4%;  Score 1016;  DB 9;  Length 290;
  Best Local Similarity   66.2%;  Pred. No. 1.1e-96;
```

```
  Matches  190;  Conservative   43;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |:||||| :|  :| |:| ::| :  | |:||||||::||||||||||| |:|||  ||
Db          2 AQIIDGKAIAADVRREVAADVAALSSAHGLVPGLAVVIVGSRKDSQTYVNMKRKACAEVG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | | :|:|||| ||  ::  :   |:||:|||||||||| |:|:| ||: :||||||||
Db         62 ICSIDVDLPEDISETALVAEVHRLNADPAVHGILVQLPLPKHINEEKILSEISIEKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:||:| |||||||||||||| | || :|||||||:||||||||:| :|||
Db        122 FHPLNIGKLAMKGREPLFVPCTPKGCMELLSRSGVTVKGKRAVVVGRSNIVGLPVSLLLL 181

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:||||:||| |  || | |:||||||| | | :::| |||||||:|||| | :|||
Db        182 KADATVSVVHSRTPDPESIVREADIVIAAAGQAMMIKGDWIKPGAAVIDVGTNSIDDPTR 241

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                | |||||| :  |||:|| :||||||||||||:|||| ||:: |||
Db        242 KSGVRLVGDVDFAAASKVAGYLTPVPGGVGPMTVAMLLKNTVDGAKR 288


RESULT 11
US-10-723-061-16
; Sequence 16, Application US/10723061
; Publication No. US20040132150A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV1
;  CURRENT APPLICATION NUMBER: US/10/723,061
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  PRIOR APPLICATION NUMBER: US 09/903,814
;  PRIOR FILING DATE: 2001-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
;   LENGTH: 290
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-723-061-16

  Query Match             68.4%;  Score 1016;  DB 16;  Length 290;
  Best Local Similarity   66.2%;  Pred. No. 1.1e-96;
  Matches  190;  Conservative   43;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |:||||| :|  :| |:| ::| :  | |:||||||::||||||||||| |:|||  ||
Db          2 AQIIDGKAIAADVRREVAADVAALSSAHGLVPGLAVVIVGSRKDSQTYVNMKRKACAEVG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | | :|:|||| ||  ::  :   |:||:|||||||||| |:|:| ||: :||||||||
Db         62 ICSIDVDLPEDISETALVAEVHRLNADPAVHGILVQLPLPKHINEEKILSEISIEKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:||:| |||||||||||||| | || :|||||||:||||||||:| :|||
Db        122 FHPLNIGKLAMKGREPLFVPCTPKGCMELLSRSGVTVKGKRAVVVGRSNIVGLPVSLLLL 181

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:||||:||| |  || | |:||||||| | | :::| |||||||:|||| | :|||
Db        182 KADATVSVVHSRTPDPESIVREADIVIAAAGQAMMIKGDWIKPGAAVIDVGTNSIDDPTR 241

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                | |||||| :  |||:|| :||||||||||||:|||| ||:: |||
Db        242 KSGVRLVGDVDFAAASKVAGYLTPVPGGVGPMTVAMLLKNTVDGAKR 288


RESULT 12
US-10-437-963-176965
; Sequence 176965, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 176965
;   LENGTH: 292
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_74664C.1.pep
US-10-437-963-176965

  Query Match             68.1%;  Score 1012;  DB 16;  Length 292;
  Best Local Similarity   66.6%;  Pred. No. 2.8e-96;
  Matches  191;  Conservative   41;  Mismatches   55;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |:||||| ||  || |:| ::| :  |  :||||||::|||||||||||: |:|||  ||
Db          2 AQIIDGKAVAADIRREVAADVAALSSAHNLVPGLAVVIVGSRKDSQTYVQMKRKACAEVG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |:| :|:| || ||  ::  :   |:||:|||||||||| |:|:| ||| :|:||||||
Db         62 IRSVDVDLAEDISEAALVAEVHRLNADPAVHGILVQLPLPKHINEEKILNEISLEKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:|||| |:|||||||||| | || | |||||||:||||||||:| :|||
Db        122 FHPLNIGKLAMKGRDPLFLPCTPKGCMELLTRSGVTINGKRAVVVGRSNIVGLPVSLLLL 181

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:|||||||| |   || | |:||||||| | | :::| ||||||:|||| | : ||
Db        182 KADATVSIVHSRTPNPESIVREADIVIAAAGQAMMIKGDWIKPGAAVIDVGTNSISDPTR 241

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                |||||||  : | ||:|| :||||||||||:|||| ||:: |||
Db        242 KSGYRLVGDVDFAEVSKVAGHLTPVPGGVGPMTVAMLLKNTVDGAKR 288


RESULT 13
US-10-424-599-268535
; Sequence 268535, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 268535
;   LENGTH: 294
;   TYPE: PRT
```

```
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_8450C.1.pep
US-10-424-599-268535

  Query Match             68.0%;  Score 1010;  DB 12;  Length 294;
  Best Local Similarity   65.5%;  Pred. No. 4.5e-96;
  Matches  188;  Conservative   47;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :|||| ||: || ||| |: ::    | ||||||::||:|||||:||  |:|||  :|
Db          2 ATVIDGKAVAQTIRSEIADEVRQLSQKYGKVPGLAVVIVGNRKDSQSYVGMKRKACAELG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||::|:|||  |: |::| :   |::| |||||||||||| |:|:| :|  :|:||||||
Db         62 IKSFDVDLPEQVSQAELIKQVHQLNANPDVHGILVQLPLPKHINEEEVLTEISLEKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:|||| |:||||| |:||| | || ||||:|||:|||||||:||:|||
Db        122 FHPLNIGKLAMKGRDPLFLPCTPKACIELLQRSGVSIKGKKAVVVGRSNIVGLPASLLLL 181

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:|||:|:|||:| :|| | |:|||||| |   :::|||||||||:|||| | ||||
Db        182 KADATVTIIHSHTSQPESIIREADIVIAAAGQPMMIKGSWIKPGAAVIDVGTNAVDDPTK 241

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
              |||||||| : ||||:|| ||||||||||||: ||| |||: |||
Db        242 KSGYRLVGDVDFGEASKVAGWITPVPGGVGPMTVTMLLKNTLDGAKR 288


RESULT 14
US-10-425-114-44200
; Sequence 44200, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 44200
;   LENGTH: 321
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700847483_FLI.pep
US-10-425-114-44200

  Query Match             66.9%;  Score 994;  DB 12;  Length 321;
  Best Local Similarity   65.2%;  Pred. No. 2.4e-94;
  Matches  187;  Conservative   45;  Mismatches   55;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :|||| ||: || ||| |:  :    | ||||||::||:|||||:||  |:|||  :|
Db         29 ATVIDGKAVAQIIRSEIADEVRLLSQKYGKVPGLAVVIVGNRKDSQSYVGMKRKACAELG 88

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||::|:|||  |: |::| :   | :| ||||||||||| |:|:| :|  :|:||||||
Db         89 IKSFDVDLPEQVSQAELIKQVHELNVNPDVHGILVQLPLPKHINEEKVLTEISLEKDVDG 148

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:|||| |:||||| |:||| | || ||||:|||:|||||||:||:|||
Db        149 FHPLNIGKLAMKGRDPLFLPCTPKACIELLQRSGVSIKGKKAVVVGRSNIVGLPASLLLL 208

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:|||:||||:| :|| |  :|||||| |   :::|:|||||||:|||| | ||||

Db        209 KADATVTIVHSHTSQPESIIHEADIVIAAAGQPMMIKGNWIKPGAAVIDVGTNAVDDPTK 268

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
              |||||||| :|||||:|| |||||||||| ||: ||| ||:| |||
Db        269 KSGYRLVGDVDFEEASKVAGWITPVPGGVGSMTVTMLLKNTVEGAKR 315


RESULT 15
US-10-437-963-125501
; Sequence 125501, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 125501
;   LENGTH: 395
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_28138C.1.pep
US-10-437-963-125501

  Query Match             66.5%;  Score 988;  DB 16;  Length 395;
  Best Local Similarity   65.0%;  Pred. No. 1.4e-93;
  Matches  186;  Conservative   45;  Mismatches   55;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | |||||  ||: || :||  |: :||:|:|  ||||||:|||  |:||::|||  |  |  ||  ||
Db         98 ATIIDGKSVAEDIRFQIAEEVRQMKNAVGHVPGLAVVLVGDRRDSESYVRYKIKGCEEVG 157

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||     || : :|| |:  :: || ||||||||||||||| ||::| ||:|:|:||||||
Db        158 IKSLLAELPGNCTEDVVVDSVSRFNEDPSVHGILVQLPLPQHMDEERILSAISLEKDVDG 217

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||:|  ||::  |  |  ||||   |  |:||| :  |:|:  ||    ||||| :||:|  :||||
Db        218 FHPLNVGNLALRSRKPLFVPCAAKACLELLLQSGIELMGKHVTVIGRSKVVGLPTSLLLQ 277

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              : :|||||:|: |  |||||||:||||:| ||||||||||:|  || :|||| ||::|| |
Db        278 RHHATVSIIHAFTTNPEEITRQSDIVISAAGVANLVRGSWLKKGAVVIDVGTNPIEDPTS 337

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAK 286
              ||||  ||||:||| |:|  ||||||||||:||||||:|||:|||
Db        338 DYGYRLTGDVCFEEAVKLASAITPVPGGVGPVTIAMLLANTLDSAK 383


Search completed: July 24, 2004, 07:09:11
Job time : 970 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 24, 2004, 06:34:25 ; Search time 57 Seconds
(without alignments)
264.470 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1486	100.0	292	4	US-09-903-814A-18	Sequence 18, Appl
2	1030	69.3	242	4	US-09-903-814A-20	Sequence 20, Appl
3	1016	68.4	290	4	US-09-903-814A-16	Sequence 16, Appl
4	712.5	47.9	284	4	US-09-107-532A-7288	Sequence 7288, Ap
5	701	47.2	289	4	US-09-903-814A-22	Sequence 22, Appl
6	684	46.0	318	4	US-09-489-039A-13257	Sequence 13257, A
7	679.5	45.7	290	4	US-09-134-001C-3401	Sequence 3401, Ap
8	677	45.6	318	4	US-09-252-991A-30283	Sequence 30283, A
9	663.5	44.7	280	4	US-09-134-000C-3865	Sequence 3865, Ap
10	661	44.5	290	4	US-09-543-681A-7078	Sequence 7078, Ap
11	623.5	42.0	344	3	US-09-318-448-21	Sequence 21, Appl
12	608.5	40.9	314	4	US-09-328-352-5492	Sequence 5492, Ap
13	518	34.9	271	4	US-09-198-452A-353	Sequence 353, App
14	104.5	7.0	482	4	US-09-252-991A-19130	Sequence 19130, A
15	99.5	6.7	231	4	US-08-426-630-54	Sequence 54, Appl
16	98	6.6	382	4	US-09-134-001C-3765	Sequence 3765, Ap
17	92	6.2	1025	4	US-09-711-164-443	Sequence 443, App
18	90.5	6.1	409	4	US-09-134-001C-3790	Sequence 3790, Ap
19	88	5.9	202	4	US-08-311-731A-13	Sequence 13, Appl
20	85.5	5.8	628	4	US-09-252-991A-32414	Sequence 32414, A
21	85.5	5.8	650	4	US-09-252-991A-27296	Sequence 27296, A
22	83.5	5.6	648	3	US-09-183-706-43	Sequence 43, Appl
23	83.5	5.6	648	4	US-09-567-995-43	Sequence 43, Appl
24	83.5	5.6	865	4	US-09-711-164-354	Sequence 354, App
25	83	5.6	300	4	US-09-328-352-6714	Sequence 6714, Ap
26	82.5	5.6	432	4	US-09-489-039A-13935	Sequence 13935, A
27	82.5	5.6	524	4	US-09-540-236-2706	Sequence 2706, Ap
28	82	5.5	502	4	US-09-134-001C-3598	Sequence 3598, Ap
29	82	5.5	549	3	US-08-886-886-13	Sequence 13, Appl
30	82	5.5	549	4	US-09-134-000C-5876	Sequence 5876, Ap
31	81.5	5.5	262	4	US-09-134-000C-4224	Sequence 4224, Ap
32	81.5	5.5	430	4	US-08-311-731A-153	Sequence 153, App
33	81.5	5.5	664	4	US-09-252-991A-30117	Sequence 30117, A
34	81	5.5	347	4	US-09-107-532A-6054	Sequence 6054, Ap
35	81	5.5	520	1	US-08-706-292-2	Sequence 2, Appli
36	81	5.5	520	3	US-09-032-365A-15	Sequence 15, Appl
37	81	5.5	904	4	US-09-543-681A-4485	Sequence 4485, Ap
38	80.5	5.4	357	3	US-09-116-115-17	Sequence 17, Appl
39	80.5	5.4	357	4	US-09-541-762-17	Sequence 17, Appl
40	80	5.4	680	4	US-09-717-364A-19	Sequence 19, Appl
41	80	5.4	924	4	US-09-267-311-2	Sequence 2, Appli
42	79.5	5.3	269	4	US-09-252-991A-27159	Sequence 27159, A
43	79.5	5.3	523	2	US-08-482-728A-19	Sequence 19, Appl
44	79.5	5.3	523	3	US-09-028-366-4	Sequence 4, Appli
45	79.5	5.3	523	4	US-09-715-285-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-903-814A-18
; Sequence 18, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB1179 USDIV
; CURRENT APPLICATION NUMBER: US/09/903,814A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US 60/092,869
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: US 09/351,703
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-903-814A-18

Query Match 100.0%; Score 1486; DB 4; Length 292;
Best Local Similarity 100.0%; Pred. No. 8.5e-165;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
```

```
RESULT 2
US-09-903-814A-20
; Sequence 20, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV
;  CURRENT APPLICATION NUMBER: US/09/903,814A
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
;   LENGTH: 242
;   TYPE: PRT
;   ORGANISM: Glycine max
US-09-903-814A-20

  Query Match             69.3%;  Score 1030;  DB 4;  Length 242;
  Best Local Similarity   80.5%;  Pred. No. 1.1e-111;
  Matches  194;  Conservative   22;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         52 KKKACEAVGIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNA 111
              :|||||:||| | | ||||:|:|:||| :|| :| |||||||||||||| |||::|||||
Db          2 EKKACESVGINSLEANLPENSTEEEVLNYIAGYNDDPSVHGILVQLPLPSHMNEQNILNA 61

Qy        112 VSIEKDVDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIV 171
              | |||||||||||||||||:||:| |||||||||:|||||| | ||||||||||||||
Db         62 VRIEKDVDGFHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNVSIKGKRAVVIGRSNIV 121

Qy        172 GMPAALLLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVG 231
              |||||||||: :|||||||| |  |||| |||||:||||| ||:|||||||||| |||||
Db        122 GMPAALLLQREDATVSIVHSRTSNPEEIIRQADIIIAAVGQANMVRGSWIKPGAVIIDVG 181

Qy        232 INPVDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKF 291
              ||||:|| ||||||:|||||||||| :||  :|||||||||||||| ||| ||||:| |
Db        182 INPVEDPNSPRGYKLVGDVCYEEAIRIASAVTPVPGGVGPMTIAMLLQNTLISAKRVHNF 241

Qy        292 K 292
              :
Db        242 E 242


RESULT 3
US-09-903-814A-16
; Sequence 16, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV
;  CURRENT APPLICATION NUMBER: US/09/903,814A
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
;   LENGTH: 290
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-903-814A-16

  Query Match             68.4%;  Score 1016;  DB 4;  Length 290;
  Best Local Similarity   66.2%;  Pred. No. 6.3e-110;
  Matches  190;  Conservative   43;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |:||||| :| :| |:| ::| :  | |:||||||::||||||||||  |:|||  ||
Db          2 AQIIDGKAIAADVRREVAADVAALSSAHGLVPGLAVVIVGSRKDSQTYVNMKRKACAEVG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | | :|:|||| ||  ::  :   |:||:|||||||||| |:|:| ||: :||||||||
Db         62 ICSIDVDLPEDISETALVAEVHRLNADPAVHGILVQLPLPKHINEEKILSEISIEKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:||:| |||||||||||||| | || :|||||||:|||||||:| :|||
Db        122 FHPLNIGKLAMKGREPLFVPCTPKGCMELLSRSGVTVKGKRAVVVGRSNIVGLPVSLLLL 181

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:||||:||| |  || | |:||||||| | | :::| |||||||:|||| | :|||
Db        182 KADATVSVVHSRTPDPESIVREADIVIAAAGQAMMIKGDWIKPGAAVIDVGTNSIDDPTR 241

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
              | |||||| :  |||:|| :||||||||||:|||| ||:: |||
Db        242 KSGVRLVGDVDFAAASKVAGYLTPVPGGVGPMTVAMLLKNTVDGAKR 288


RESULT 4
US-09-107-532A-7288
; Sequence 7288, Application US/09107532A
; Patent No. 6583275
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                           ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 7310
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: PC
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/085,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;  INFORMATION FOR SEQ ID NO: 7288:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 284 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
```

```
;       ORIGINAL SOURCE:
;            ORGANISM: Enterococcus faecium
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...284
;       SEQUENCE DESCRIPTION: SEQ ID NO: 7288:
US-09-107-532A-7288

  Query Match             47.9%;  Score 712.5;  DB 4;  Length 284;
  Best Local Similarity   51.0%;  Pred. No. 1.7e-74;
  Matches  146;  Conservative   54;  Mismatches   77;  Indels    9;  Gaps    5;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |::|:|| :|:::: ||| :| |:||  |: ||| |:|||    || |||||::  : :|
Db          2 AELINGKELAEKMQAEIAEKIQKLKDN-GIHPGLVVLLVGENPASQIYVRNKERVAKEIG 60

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | |      ||  ||||:|  |  :| ||| |||||||||| |:::| || |:  ||||||
Db         61 IHSLVERYPETISEDELLAEIEKYNQDPSFHGILVQLPLPKHIDEEKILLAIRPEKDVDG 120

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||:|:||| : |: |  :|||| | |::   | ::  |||||||||||||| | | ||
Db        121 FHPMNLGRL-LAGK-PDKIPCTPYGIMKMFEAYSIDPAGKRAVVIGRSNIVGKPMAQLLL 178

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              |:|||:: || |:   |: ::|||:: |:|  : |   ::|||| :||||:|
Db        179 MADATVTVAHSKTENLAELAQEADILVVAIGRGHFVTKEFVKPGAVVIDVGMN-----RD 233

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAK 286
              | :|:||| ::|  ||| ||||| ||||||| ||:  |::|:
Db        234 ANG-KLIGDVKFDEVEPIAGWITPVPKGVGPMTITMLMYQTVKSAE 278


RESULT 5
US-09-903-814A-22
; Sequence 22, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl
;  APPLICANT:  Famodu, Layo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Thorpe, Cathy
;  TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
;  FILE REFERENCE: BB1179 USDIV
;  CURRENT APPLICATION NUMBER: US/09/903,814A
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: US 60/092,869
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: US 09/351,703
;  PRIOR FILING DATE: 1999-07-12
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
;   LENGTH: 289
;   TYPE: PRT
;   ORGANISM: Triticum aestivum
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (195)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (227)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (233)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (235)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (252)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (256)..(257)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (260)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (267)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (272)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (278)..(279)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (281)..(282)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (286)
;   OTHER INFORMATION: Xaa can be any naturally ocurring amino acid
US-09-903-814A-22

  Query Match             47.2%;  Score 701;  DB 4;  Length 289;
  Best Local Similarity   58.4%;  Pred. No. 3.8e-73;
  Matches  135;  Conservative   37;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |:||||| :| :|| ||  |:| :  |  :||||||::||||||||||: |:|||  ||
Db          7 AQIIDGKAIAAEIRREIGAEVAVLSSAHNIVPGLAVVIVGSRKDSQTYVQMKRKACAEVG 66

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |:|::|:|||| :|  ::  :  |:||:||||||||||| |:|:|||| :||||||||
Db         67 IRSFDVDLPEDIAEAALVAEVHRLNADPAVHGILVQLPLPKHINEENILNQISIEKDVDG 126

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:|||| |||||||||||||| | || :||| |||:|||||||:|:
Db        127 FHPLNIGKLAMKGRDPLFVPCTPKGCMELLSRSGVTVKGKHAVVVGRSNIVGLPSIPSPS 186

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVG 231
              ::         |    |: |: : |||||| | | :::| | |      | ||
Db        187 ESGRYRVDXASTDPNPQTISVKQDIVIAAAGQAMMIKGDWXKQKRNXIXVG 237


RESULT 6
US-09-489-039A-13257
; Sequence 13257, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 13257
;   LENGTH: 318
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
```

US-09-489-039A-13257

Query Match 46.0%; Score 684; DB 4; Length 318;
Best Local Similarity 49.0%; Pred. No. 4.3e-71;
Matches 142; Conservative 48; Mismatches 90; Indels 10; Gaps 4;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGV-VPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              ||||||| :|:|:| |:| | | : | |   |||||:|||   || || :|:|||| |
Db         33 AKIIDGKTIAQQVRSEVA-EKVKARVAAGKRAPGLAVVLVGSNPASQIYVGSKRKACEEV 91

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVD 119
              |  |   :||| :|| |:|: | | |:| :: ||||||||  :::  :|  :: :||||
Db         92 GFVSRSYDLPETTSEAELLELIDTLNADKTIDGILVQLPLPAGIDNVKVLERIAPDKDVD 151

Qy        120 GFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLL 179
              |||| |:|||    | |   ||||:| : || || ::  |  ||||| ||||| | :: |
Db        152 GFHPYNVGRLCQ--RAPRLRPCTPRGIVTLLERYNIDTYGLNAVVIGASNIVGRPMSMEL 209

Qy        180 QKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPE 239
                | :: |  ||         ||:| |||    : | ||| || ::||||| ::
Db        210 LLAGCTTTVTHRFTKNLRHHVENADLLIVAVGKPGFIPGEWIKEGAIVVDVGINRLESG- 268

Qy        240 SPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIH 289
                ::|||| ||:|:: |  |||||||||||||:| |: |||:: :  |
Db        269 -----KVVGDVVYEDAAERASYITPVPGGVGPMTVATLIQNTLQACEEYH 313
```

RESULT 7
US-09-134-001C-3401
; Sequence 3401, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3401
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3401

Query Match 45.7%; Score 679.5; DB 4; Length 290;
Best Local Similarity 48.6%; Pred. No. 1.2e-70;
Matches 139; Conservative 59; Mismatches 79; Indels 9; Gaps 3;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |||:||| :||: |: :  ::  :|:  |  | |:|||||:   ||:||::|||| | :|
Db          7 AKILDGKQIAKEYRQRLKNQVNDLKE-YGFTPKLSVILVGNDGASQSYVKSKKKAAEKIG 65

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              : |  ::| | :||: ||  :   |:| :| |||||:|||  :::: :| |:: ||||||
Db         66 MISEIIHLDESTSEEVVLSELNRLNNDDTVSGILVQVPLPKQVSEQKVLEAINPEKDVDG 125

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||:|||:| :   :  ||||||  | ||:|    : ::|| |||||||:||| | : ||
Db        126 FHPINIGKLYID--EQTFVPCTPLGIMEILKHADINLEGKNAVVIGRSHIVGQPVSKLLL 183

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              :|||||:|:|| ||      :|||::::|||   ||    :| || |||||  | ::
Db        184 QANATVTILHSRTKNMNAHLKQADVIVSAVGQPGLVTKENVKKGAVIIDVGNTPDENG-- 241

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAK 286
                :| ||| |:|  :||  |||||||||:|| |:|:||| : |
Db        242 ----KLKGDVAYDEVKEIASAITPVPGGVGPLTITMVLNNTLLAEK 283
```

RESULT 8
US-09-252-991A-30283
; Sequence 30283, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30283
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30283

Query Match 45.6%; Score 677; DB 4; Length 318;
Best Local Similarity 49.8%; Pred. No. 2.8e-70;
Matches 144; Conservative 42; Mismatches 95; Indels 8; Gaps 2;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |::|||| :|  :|::||  : : :     ||||||||||:   || || :|:| || ||
Db         37 AQLIDGKAIAANLRQQIAQRVTERRQQGLRVPGLAVILVGTDPASQVYVAHKRKDCEEVG 96

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
                |  :|| ::|:|::|  |   | ||:: |||||||||| |::  :|  :  :|||||
Db         97 FLSQAYDLPAETSQDDLLALIDRLNDDPAIDGILVQLPLPAHLDASLLLERIHPDKDVDG 156

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              ||| ||||||   | |   ||||||  | ||   | :: |  |||:| ||||| | || |
Db        157 FHPYNIGRLAQ--RMPLLRPCTPKGIMTLLASTGADLYGMDAVVVGASNIVGRPMALELL 214

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
                ||:: |  |:   :   :||:|: | |   ||:| ||| || :||||||   |
Db        215 LGGCTVTVTHRFTRDLADHVSRADLVVVAAGKPGLVKGEWIKEGAIVIDVGINRQADG-- 272

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIH 289
                |||||| || |:: |  |||||||||||| | || ||| :|: :|
Db        273 ----RLVGDVEYEVAAQRASWITPVPGGVGPMTRACLLENTLHAAEHLH 317
```

RESULT 9
US-09-134-000C-3865
; Sequence 3865, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3865
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-3865

Query Match 44.7%; Score 663.5; DB 4; Length 280;
Best Local Similarity 50.0%; Pred. No. 8.6e-69;
Matches 135; Conservative 50; Mismatches 76; Indels 9; Gaps 5;

```
Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
              :|:|: :| |:: ||  :: ||    |: ||| |:|||    |||||||::|   :||
Db         12 VINGRELADQMQAEIQKDVEKMTQQ-GIQPGLVVLLVGENPASQTYVRNKERAAAKIGIL 70

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
              |    |||  ||:|:|  |  :| |  ||||||||| |:::| || |:  :|||||||
Db         71 SKVEKLPETISEEELLAEIDKYNQDSRFHGILVQLPLPKHIDEEKILLAIDPKKDVDGFH 130

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
              |:|:||| : |: |  :|||| | |::   | ::: |||||||||||||| | | ||
Db        131 PMNLGRLFV-GK-PEMIPCTPYGIMKMFEAYDIDLTGKRAVVIGRSNIVGKPMAQLLLMK 188

Qy        183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
              ||||:| || |:   |: ::|||:: |:|  : |   ::|||| :||||:|      :
Db        189 NATVTIAHSKTEHLAEVAKEADILVVAIGRGHFVTKEFVKPGAVVIDVGMN-----RNQE 243

Qy        243 GYRLVGDVCYEEASKIAGLITPVPGGVGPM 272
              | :|:||| ::| |:||  ||||| |||||
Db        244 G-KLIGDVAFDEVSEIASYITPVPKGVGPM 272


RESULT 10
US-09-543-681A-7078
; Sequence 7078, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7078
;   LENGTH: 290
;   TYPE: PRT
;   ORGANISM: Proteus mirabilis
US-09-543-681A-7078

  Query Match             44.5%;  Score 661;  DB 4;  Length 290;
  Best Local Similarity   48.3%;  Pred. No. 1.8e-68;
  Matches  140;  Conservative   47;  Mismatches   93;  Indels   10;  Gaps    4;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGV-VPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              |:|||| :|: || |:| |  | :  ||   |||||||||    || || :|:|||: |
Db          3 ARIIDGKTIARTIRSEVA-EKVKQRIKIGKRAPGLAVILVGDNPASQIYVASKRKACDEV 61

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVD 119
              | |  :||: :|| ::|  | | | | :: |||||||||  :::  :|  :  :||||
Db         62 GFISRSYDLPDTTSEADLLNLIDTLNEDNTIDGILVQLPLPAGIDNVKVLERIHPDKDVD 121

Qy        120 GFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLL 179
              |||| |||||    | |   ||||:| : || |  : : |  ||:|| ||||| | :| |
Db        122 GFHPYNIGRLCQ--RAPKLRPCTPRGIVTLLERCNIPMNGLNAVIIGASNIVGRPMSLEL 179

Qy        180 QKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPE 239
                |  | :: |  ||         ||:|: |||  | : | ||||||| :||||||| :::
Db        180 LLAGCTTTVTHRFTKDLRFHVEHADLVVVAVGKPNFIPGEWIKPGAIVIDVGINRLENG- 238

Qy        240 SPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIH 289
                  ::|||| : :||: || |:|||||||||:| |: |||:: :  |
Db        239 -----KVVGDVDFAQASQRAGWISPVPGGVGPMTVATLIQNTLQACEEYH 283


RESULT 11
US-09-318-448-21
; Sequence 21, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
;  APPLICANT: Johnson, William G.
;  APPLICANT: Stenroos, Edward S.
;  TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
;  FILE REFERENCE: 601-1-057
;  CURRENT APPLICATION NUMBER: US/09/318,448
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 344
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-318-448-21

  Query Match             42.0%;  Score 623.5;  DB 3;  Length 344;
  Best Local Similarity   46.2%;  Pred. No. 5.6e-64;
  Matches  138;  Conservative   52;  Mismatches   94;  Indels   15;  Gaps    5;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :|  |: :|:||::|:  |: :   :    | |:|||||     | :|| || :|   ||
Db         31 AVVISGRKLAQQIKQEVRQEVEEWVASGNKRPHLSVILVGENPASHSYVLNKTRAAAVVG 90

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | |  :  |   ||:|:|  |    |:| :| |:|||||| |:::  | |||| :|||||
Db         91 INSETIMKPASISEEELLNLINKLNNDDNVDGLLVQLPLPEHIDERRICNAVSPDKDVDG 150

Qy        121 FHPLNIGRLAMQGRDPF-FVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLL 179
              || :|:||: :   | :  :| || |  |:: | |:   ||  || |||  |||| |:||
Db        151 FHVINVGRMCL---DQYSMLPATPWGVWEIIKRTGIPTLGKNVVVAGRSKNVGMPIAMLL 207

Qy        180 QK--------ANATVSIVHSNTKKPE--EITRQADIVIAAVGVANLVRGSWIKPGAAIID 229
                        :|||:| |  | | :  : |  |||||:| |: ||:    || |||:||
Db        208 HTDGAHERPGGDATVTISHRYTPKEQLKKHTILADIVISAAGIPNLITADMIKEGAAVID 267

Qy        230 VGINPVDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||| | || : :  :|||||  :|   :  || |||||||||||||:|||: ||: :||::
Db        268 VGINRVHDPVTAKP-KLVGDVDFEGVRQKAGYITPVPGGVGPMTVAMLMKNTIIAAKKV 325


RESULT 12
US-09-328-352-5492
; Sequence 5492, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5492
;   LENGTH: 314
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-5492

  Query Match             40.9%;  Score 608.5;  DB 4;  Length 314;
  Best Local Similarity   43.2%;  Pred. No. 2.7e-62;
  Matches  124;  Conservative   56;  Mismatches   98;  Indels    9;  Gaps    2;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | ::||: :|||| | : | :  :|    |  | || ||||    | |||| |  ||  ||
Db         34 ALVLDGRALAKQIEENLLVRVEALKAKTGRTPILATILVGDDGASATYVRMKGNACRRVG 93

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              : | :: ||:::: :::|  |    |::| ||||||:| |:|  :::    :|:|: |||||
Db         94 MDSLKIELPQETTTEQLLAEIEKLNANPDVHGILLQHPVPAQIDERACFDAISLAKDVDG 153

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
                |  ||:||   :  :   || | | :|    :|| || |||:||| |:| | |::|
Db        154 VTCLGFGRMAM--GEAAYGSATPAGIMTILKENNIEIAGKHAVVVGRSAILGKPMAMMLL 211

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
```

```
Db        212 QANATVTICHSRTQNLPELVKQADIIVGAVGKAELIQKDWIKQGAVVVDAGFHPRDGGG- 270

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
Db        271 ------VGDIQLQGIEEIASAYTPVPGGVGPMTITTLIRQTVEAAEK 311


RESULT 13
US-09-198-452A-353
; Sequence 353, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 353
;   LENGTH: 271
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
US-09-198-452A-353

  Query Match             34.9%;  Score 518;  DB 4;  Length 271;
  Best Local Similarity   41.5%;  Pred. No. 7.9e-52;
  Matches  113;  Conservative   54;  Mismatches   87;  Indels   18;  Gaps    6;

Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
Db          5 LLRGIPAAEKILQRLKEEISQSPTS----PGLAVVLIGNDPASEVYVGMKVKKATEIGII 60

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
Db         61 SKAHKLPSDSTLSSVLKLIERLNQDPSIHGILVQLPLPKHLDSEVILQAISPDKDVDGLH 120

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMP-AALLLQK 181
Db        121 PVNMGKLLL-GNFDGLLPCTPAGIIELLNYYEIPLRGRHAAIVGRSNIVGKPLAALMMQK 179

Qy        182 ---ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDP 238
Db        180 HPQTNCTVTVLHSQSENLPEILKTADIIIAALGAPLFIKETMVAPHAVIVDVGTTRV-PA 238

Qy        239 ESPRGYRLVGDV--------CYEEASKIAGLI 262
Db        239 DNAKGYTLLGDVDFNNVVTKCAESLQFLEALV 270


RESULT 14
US-09-252-991A-19130
; Sequence 19130, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 19130
;   LENGTH: 482
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19130

  Query Match              7.0%;  Score 104.5;  DB 4;  Length 482;
  Best Local Similarity   22.1%;  Pred. No. 0.0039;
  Matches   54;  Conservative   39;  Mismatches   72;  Indels   79;  Gaps   10;

Qy         29 GVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIKSYEVNLPEDSS--------EDEVL-- 78
Db         25 GLAGGVAQVL-----HAQAQVASRSHQFDAVAALEAVLHLHLGQSVAAFAGEFADELLAL 79

Qy         79 -KHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFHPLNIGRLAMQGRDPF 137
Db         80 GQRMA----------VLVVLVLVVHVAAQEALGVAAVGEGLHGVHQAAVERLA--GR--- 124

Qy        138 FVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKANATVSIVHSNTKKPE 197
Db        125 ----------RVVDRLAVDLGGARHVV------VGFGAALDLQR-------VHADPRQPL 161

Qy        198 EITRQADIV-IAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPRGYRLVGDVCYEEAS 256
Db        162 DVLHGAEVLGIHDVGAVFVLEG----------------------GHRFAGTVAFLEEE 197

Qy        257 KIAG 260
Db        198 HLLG 201


RESULT 15
US-08-426-630-54
; Sequence 54, Application US/08426630
; Patent No. 6656709
;  GENERAL INFORMATION:
;    APPLICANT:  BLANCHE, FRANCIS; CAMERON, BEATRICE; CROUZET,
;    APPLICANT:  JOEL;DEBUSSCHE, LAURENT; LEVCY SCHIL, SOPHIE;
;    APPLICANT:  THIBAUT, DENIS
;    TITLE OF INVENTION:  POLYPEPTIDES INVOLVED IN THE
;    TITLE OF INVENTION:  BIOSYNTHESIS OF COBALAMINS AND/OR COBAMIDES, DNA SEQUENCES
;    TITLE OF INVENTION:  CODING FOR THESE POLYPEPTIDES, PREPARATION METHOD AND THEIR
;    TITLE OF INVENTION:  USE.
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  555 13TH STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  DISTRICT OF COLUMBIA
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/426,630
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/916,151
;      FILING DATE:  14-SEP-1992
;      APPLICATION NUMBER:  PCT/FR91/00054
;      FILING DATE:  30-JAN-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  F. F. CALVETTI
;      REGISTRATION NUMBER:  28,557
;      REFERENCE/DOCKET NUMBER:  1290-7213
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 857-7887
;      TELEFAX:  (202) 857-7929
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  231 amino acids
;      TYPE:  Amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  Linear
```

; MOLECULE TYPE: protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Methanobacterium ivanovii
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; FEATURE:
; NAME/KEY: CORA
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: Translation product of SEQ ID NO:53
US-08-426-630-54

Query Match 6.7%; Score 99.5; DB 4; Length 231;
Best Local Similarity 21.3%; Pred. No. 0.0044;
Matches 56; Conservative 46; Mismatches 94; Indels 67; Gaps 12;

```
Qy          34 LAVILVGSRKDSQTYVRNKKKACEAVGI-KSYEVNLPEDSSEDEVLKHIATFNSDPSVHG 92
               :  | |||:       :  |     || : |  :| | :  : :|:||:
Db           1 MVVYLVGAGPGDPELITLK-----AVNVLKKADVVLYDKPANEEILKYA------EGAKL 49

Qy          93 ILVQLPLPHHMNDENILNAVSIEKDVDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLH- 151
               | |      ||   :| :| : :|:  :      |   : ::| |||      :|  |:|
Db          50 IYVGKQAGHHYKSQNEINTLLVEEAKE-----NDLVVRLKGGDPFVF---GRGGEEILAL 101

Qy         152 -RYGVEIKGKRAVVIGRSNIVGMPAAL---LLQKANATVSIVHSNTKKPEEITRQ----- 202
                |:: :    :| | :: :|:|  :   :  :  ||   | :  : | :  :|
Db         102 VEEGIDFE----LVPGVTSAIGVPTTIGLPVTHRGVATSFTVVTGHEDPTKCKKQVGWDF 157

Qy         203 -ADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPRGYRLVGDVCYEEASKIAGL 261
                || ::  :|: ||   :           |   |||:|        ||  |
Db         158 KADTIVILMGIGNLAENT----------AEIMKHKDPETP--------VCVIEN------ 193

Qy         262 ITPVPGGVGPMTIAMLLSNTLES 284
                       | |    ::: |||:
Db         194 --------GTMEGQRIITGTLEN 208
```

Search completed: July 24, 2004, 06:41:22
Job time : 59 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 24, 2004, 04:10:50 ; Search time 101 Seconds
(without alignments)
816.870 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1238	83.3	357	3	AAG29448	Aag29448 Arabidops
2	1238	83.3	360	3	AAG29447	Aag29447 Arabidops
3	1238	83.3	426	3	AAG29446	Aag29446 Arabidops
4	1035	69.7	299	3	AAG22173	Aag22173 Arabidops
5	1035	69.7	299	3	AAG08973	Aag08973 Arabidops
6	1017	68.4	350	3	AAG36762	Aag36762 Arabidops
7	1017	68.4	352	3	AAG36761	Aag36761 Arabidops
8	1017	68.4	498	3	AAG36760	Aag36760 Arabidops
9	1012	68.1	350	3	AAG14693	Aag14693 Arabidops
10	1012	68.1	352	3	AAG14692	Aag14692 Arabidops
11	1012	68.1	377	3	AAG14691	Aag14691 Arabidops
12	762	51.3	934	4	ABB60394	Abb60394 Drosophil
13	755.5	50.8	284	5	ABB48308	Abb48308 Listeria
14	743.5	50.0	286	5	ABP25547	Abp25547 Streptoco
15	739.5	49.8	946	6	ABR52676	Abr52676 Protein s
16	739.5	49.8	946	6	ABP57428	Abp57428 Saccharom
17	735.5	49.5	938	7	ADB76851	Adb76851 A. gossyp
18	735	49.5	816	4	ABG02709	Abg02709 Novel hum
19	724	48.7	285	6	ABU01214	Abu01214 S. pneumo
20	724	48.7	286	3	AAY81531	Aay81531 Streptoco
21	721	48.5	284	5	ABP25546	Abp25546 Streptoco
22	713.5	48.0	975	6	ABR52679	Abr52679 Protein s
23	712.5	47.9	284	7	ADC97661	Adc97661 E. faeciu
24	707.5	47.6	294	5	ABB54169	Abb54169 Lactococc
25	679.5	45.7	290	5	ABP38556	Abp38556 Staphyloc
26	678	45.6	262	6	ABP77006	Abp77006 N. gonorr
27	675	45.4	301	6	ABM70041	Abm70041 Photorhab
28	670.5	45.1	294	6	ABM73307	Abm73307 Staphyloc
29	652.5	43.9	303	4	ABB66599	Abb66599 Drosophil
30	623.5	42.0	344	4	AAB50287	Aab50287 Human sch
31	623.5	42.0	344	6	ABR47536	Abr47536 Breast ca
32	623.5	42.0	359	4	AAM79919	Aam79919 Human pro
33	623.5	42.0	384	4	AAM78935	Aam78935 Human pro
34	614.5	41.4	388	4	ABG20483	Abg20483 Novel hum
35	608.5	40.9	314	6	ADA34205	Ada34205 Acinetoba
36	604.5	40.7	289	6	AAO30964	Aao30964 Human tet
37	602	40.5	292	2	AAW98386	Aaw98386 H. pylori
38	592	39.8	161	3	AAG22174	Aag22174 Arabidops
39	592	39.8	161	3	AAG08974	Aag08974 Arabidops
40	577.5	38.9	292	5	ABP65840	Abp65840 Bifidobac
41	576.5	38.8	295	2	AAY37134	Aay37134 Protein i
42	574.5	38.7	284	4	AAB79976	Aab79976 Corynebac
43	574.5	38.7	284	4	AAG90467	Aag90467 C glutami
44	574	38.6	219	2	AAW60962	Aaw60962 Streptoco
45	574	38.6	298	4	AAU48214	Aau48214 Propionib

ALIGNMENTS

RESULT 1
AAG29448
ID AAG29448 standard; protein; 357 AA.
XX
AC AAG29448;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35041.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.

PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 83.3%; Score 1238; DB 3; Length 357;
Best Local Similarity 79.1%; Pred. No. 1.9e-123;
Matches 231; Conservative 36; Mismatches 25; Indels 0; Gaps 0;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :|||| |||:||:|| :|:::||::|||:|||||||||| |||| ||||||||||::||
Db         66 AIVIDGKAVAKKIRDEITIEVSRMKESIGVIPGLAVILVGDRKDSATYVRNKKKACDSVG 125

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||:|| | |||||:|||| :: || |||||||||||||| ||:::|||||||||||||
Db        126 IKSFEVRLAEDSSEEEVLKSVSGFNDDPSVHGILVQLPLPSHMDEQNILNAVSIEKDVDG 185

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||||||:||:| |||||||||:|||||| :|||||||||||||||||||||||||
Db        186 FHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNIEIKGKRAVVIGRSNIVGMPAALLLQ 245

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              : :|||||:|| || ||||||:|||:|:||| |:|||||||| :||||||||:|| :
Db        246 REDATVSIIHSRTKNPEEITREADIIISAVGQPNMVRGSWIKPGAVLIDVGINPVEDPSA 305

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
              ||||||||:|||||||:|  ||||||||||||||||||| |||||| |:
Db        306 ARGYRLVGDICYEEASKVASAITPVPGGVGPMTIAMLLSNTLTSAKRIHNFQ 357
```

RESULT 2
AAG29447
ID AAG29447 standard; protein; 360 AA.
XX
AC AAG29447;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35040.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.

PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 83.3%; Score 1238; DB 3; Length 360;
Best Local Similarity 79.1%; Pred. No. 1.9e-123;
Matches 231; Conservative 36; Mismatches 25; Indels 0; Gaps 0;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :|||| |||:||:|| :|:::||::|||:||||||| |||| ||||||||||::||
Db         69 AIVIDGKAVAKKIRDEITIEVSRMKESIGVIPGLAVILVGDRKDSATYVRNKKKACDSVG 128

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||:|| | |||||:|||| :: || |||||||||||||| ||:::|||||||||||||
Db        129 IKSFEVRLAEDSSEEEVLKSVSGFNDDPSVHGILVQLPLPSHMDEQNILNAVSIEKDVDG 188

Qy        121 FHPLNIGRLAMQGRDPPFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              ||||||||||:||:| |||||||||:|||||| :||||||||||||||||||||||||||
Db        189 FHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNIEIKGKRAVVIGRSNIVGMPAALLLQ 248

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              : :|||||:|| || ||||||:|||:|:||| |:|||||||||| :|||||||:|| :
Db        249 REDATVSIIHSRTKNPEEITREADIIISAVGQPNMVRGSWIKPGAVLIDVGINPVEDPSA 308

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
              ||||||||:|||||||:|  |||||||||||||||||||| |||||| |:
Db        309 ARGYRLVGDICYEEASKVASAITPVPGGVGPMTIAMLLSNTLTSAKRIHNFQ 360
```

RESULT 3
AAG29446
ID AAG29446 standard; protein; 426 AA.
XX
AC AAG29446;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35039.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.

PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 83.3%; Score 1238; DB 3; Length 426;
Best Local Similarity 79.1%; Pred. No. 2.4e-123;
Matches 231; Conservative 36; Mismatches 25; Indels 0; Gaps 0;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :|||| |||:||:|| :|:::||::|||:|||||||| |||| ||||||||||::||
Db        135 AIVIDGKAVAKKIRDEITIEVSRMKESIGVIPGLAVILVGDRKDSATYVRNKKKACDSVG 194

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||:|| | |||||:|||| :: || |||||||||||||| ||:::|||||||||||||
Db        195 IKSFEVRLAEDSSEEEVLKSVSGFNDDPSVHGILVQLPLPSHMDEQNILNAVSIEKDVDG 254

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||||||:||:| |||||||||:|||||| :||||||||||||||||||||||||
Db        255 FHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNIEIKGKRAVVIGRSNIVGMPAALLLQ 314

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              : :|||||:|| || ||||||:|||:|:|||  |:||||||||| :||||||||:|| :
Db        315 REDATVSIIHSRTKNPEEITREADIIISAVGQPNMVRGSWIKPGAVLIDVGINPVEDPSA 374

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
              ||||||||:|||||||:|  |||||||||||||||||||| |||||| |:
Db        375 ARGYRLVGDICYEEASKVASAITPVPGGVGPMTIAMLLSNTLTSAKRIHNFQ 426
```

RESULT 4
AAG22173
ID AAG22173 standard; protein; 299 AA.
XX
AC AAG22173;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 25001.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.

XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.

PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 69.7%; Score 1035; DB 3; Length 299;
Best Local Similarity 67.4%; Pred. No. 7.8e-102;
Matches 194; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |||||| :|  || ||| |:  : :  | ||||||::||||||||||  |:|||  ||
Db          9 AKIIDGKAIAHTIRSEIAEEVRGLSEKHGKVPGLAVVIVGSRKDSQTYVNTKRKACAEVG 68

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||::| |||: || :::  :   ||:| |||||||||| |:|:|||| |:||:|||||
Db         69 IKSFDVGLPEEVSEADLISKVHELNSNPDVHGILVQLPLPKHINEENILGAISIDKDVDG 128

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:||:| |:||||||||:||| | ||:|||:||||:||||||||:| :|||
Db        129 FHPLNIGKLAMKGREPLFLPCTPKGCLELLARSGVKIKGQRAVVVGRSNIVGLPVSLLLL 188

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:|||: |||:|| || | |:|||||| | |::::|:|||||||:|||| | | ||
Db        189 KADATVTTVHSHTKDPEAIIREADIVIAACGQAHMIKGNWIKPGAAVIDVGTNAVSDPSK 248

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||||||| : ||||:|| |||||||||||||:|||| ||:: |||:
Db        249 KSGYRLVGDVDFAEASKVAGFITPVPGGVGPMTVAMLLRNTVDGAKRV 296
```

RESULT 5
AAG08973
ID AAG08973 standard; protein; 299 AA.
XX
AC AAG08973;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6721.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.

```
PR   10-SEP-1999;   99US-0153070P.
PR   13-SEP-1999;   99US-0153758P.
PR   15-SEP-1999;   99US-0154018P.
PR   16-SEP-1999;   99US-0154039P.
PR   20-SEP-1999;   99US-0154779P.
PR   22-SEP-1999;   99US-0155139P.
PR   23-SEP-1999;   99US-0155486P.
PR   24-SEP-1999;   99US-0155659P.
PR   28-SEP-1999;   99US-0156458P.
PR   29-SEP-1999;   99US-0156596P.
PR   04-OCT-1999;   99US-0157117P.
PR   05-OCT-1999;   99US-0157753P.
PR   06-OCT-1999;   99US-0157865P.
PR   07-OCT-1999;   99US-0158029P.
PR   08-OCT-1999;   99US-0158232P.
PR   12-OCT-1999;   99US-0158369P.
PR   13-OCT-1999;   99US-0159293P.
PR   13-OCT-1999;   99US-0159294P.
PR   13-OCT-1999;   99US-0159295P.
PR   14-OCT-1999;   99US-0159329P.
PR   14-OCT-1999;   99US-0159330P.
PR   14-OCT-1999;   99US-0159331P.
PR   14-OCT-1999;   99US-0159637P.
PR   14-OCT-1999;   99US-0159638P.
PR   18-OCT-1999;   99US-0159584P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   25-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.

  Query Match             69.7%;  Score 1035;  DB 3;  Length 299;
  Best Local Similarity   67.4%;  Pred. No. 7.8e-102;
  Matches  194;  Conservative   45;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              ||||||| :|  || ||| |:  : :  | ||||||::||||||||||  |:|||  ||
Db          9 AKIIDGKAIAHTIRSEIAEEVRGLSEKHGKVPGLAVVIVGSRKDSQTYVNTKRKACAEVG 68

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||::| |||: || :::  :   ||:| |||||||||||| |:|:|||| |:||:|||||
Db         69 IKSFDVGLPEEVSEADLISKVHELNSNPDVHGILVQLPLPKHINEENILGAISIDKDVDG 128

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:||:| |:|||||||:||| | ||:|||:||||:||||||||:| :|||
Db        129 FHPLNIGKLAMKGREPLFLPCTPKGCLELLARSGVKIKGQRAVVVGRSNIVGLPVSLLLL 188

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:|||: |||:|| || | |:|||||| | |::::|:|||||||:|||| | | ||
Db        189 KADATVTTVHSHTKDPEAIIREADIVIAACGQAHMIKGNWIKPGAAVIDVGTNAVSDPSK 248

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||||||| : ||||:|| |||||||||||||:|||| ||:: |||:
Db        249 KSGYRLVGDVDFAEASKVAGFITPVPGGVGPMTVAMLLRNTVDGAKRV 296


RESULT 6
AAG36762
ID   AAG36762 standard; protein; 350 AA.
XX
AC   AAG36762;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 45098.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   14-MAY-1999;   99US-0134370P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137528P.
PR   04-JUN-1999;   99US-0137502P.
PR   07-JUN-1999;   99US-0137724P.
PR   08-JUN-1999;   99US-0138094P.
PR   10-JUN-1999;   99US-0138540P.
PR   10-JUN-1999;   99US-0138847P.
PR   14-JUN-1999;   99US-0139119P.
PR   16-JUN-1999;   99US-0139452P.
PR   16-JUN-1999;   99US-0139453P.
PR   17-JUN-1999;   99US-0139492P.
PR   18-JUN-1999;   99US-0139454P.
PR   18-JUN-1999;   99US-0139455P.
PR   18-JUN-1999;   99US-0139456P.
PR   18-JUN-1999;   99US-0139457P.
PR   18-JUN-1999;   99US-0139458P.
PR   18-JUN-1999;   99US-0139459P.
PR   18-JUN-1999;   99US-0139460P.
PR   18-JUN-1999;   99US-0139461P.
```

PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.4%; Score 1017; DB 3; Length 350;
Best Local Similarity 67.5%; Pred. No. 8.4e-100;
Matches 193; Conservative 36; Mismatches 57; Indels 0; Gaps 0;

```
Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
              :||| ::|::|| :|  |: ||| |:| |||||:||| ::|||||||| |||| |||
Db         63 VIDGNVIAEEIRTKIISEVGKMKKAVGKVPGLAVVLVGEQRDSQTYVRNKIKACEETGIK 122

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
              |    |||| :| ::: :  || | |:|||||||||| |:|:  ||| | :||||||||
Db        123 SVLAELPEDCTEGQIISVLRKFNEDTSIHGILVQLPLPQHLNESKILNMVRLEKDVDGFH 182

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
              |||:| |||:||:| || ||||||:||| | |||| || ||||||||||||:| :||||:
Db        183 PLNVGNLAMRGREPLFVSCTPKGCVELLIRTGVEIAGKNAVVIGRSNIVGLPMSLLLQRH 242

Qy        183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
              :|||| ||: || || |||:||||||| |: |||||||:|||| :|||| ||:|
Db        243 DATVSTVHAFTKDPEHITRKADIVIAAAGIPNLVRGSWLKPGAVVIDVGTTPVEDSSCEF 302
```

```
Qy        243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              ||||||||||||| :|  |||||||||||||| ||| ||||:||||
Db        303 GYRLVGDVCYEEALGVASAITPVPGGVGPMTITMLLCNTLEAAKRI 348
```

RESULT 7
AAG36761
ID AAG36761 standard; protein; 352 AA.
XX
AC AAG36761;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45097.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.

```
PR   13-AUG-1999;   99US-0148684P.
PR   16-AUG-1999;   99US-0149368P.
PR   17-AUG-1999;   99US-0149175P.
PR   18-AUG-1999;   99US-0149426P.
PR   20-AUG-1999;   99US-0149722P.
PR   20-AUG-1999;   99US-0149723P.
PR   20-AUG-1999;   99US-0149929P.
PR   23-AUG-1999;   99US-0149902P.
PR   23-AUG-1999;   99US-0149930P.
PR   25-AUG-1999;   99US-0150566P.
PR   26-AUG-1999;   99US-0150884P.
PR   27-AUG-1999;   99US-0151065P.
PR   27-AUG-1999;   99US-0151066P.
PR   27-AUG-1999;   99US-0151080P.
PR   30-AUG-1999;   99US-0151303P.
PR   31-AUG-1999;   99US-0151438P.
PR   01-SEP-1999;   99US-0151930P.
PR   07-SEP-1999;   99US-0152363P.
PR   10-SEP-1999;   99US-0153070P.
PR   13-SEP-1999;   99US-0153758P.
PR   15-SEP-1999;   99US-0154018P.
PR   16-SEP-1999;   99US-0154039P.
PR   20-SEP-1999;   99US-0154779P.
PR   22-SEP-1999;   99US-0155139P.
PR   23-SEP-1999;   99US-0155486P.
PR   24-SEP-1999;   99US-0155659P.
PR   28-SEP-1999;   99US-0156458P.
PR   29-SEP-1999;   99US-0156596P.
PR   04-OCT-1999;   99US-0157117P.
PR   05-OCT-1999;   99US-0157753P.
PR   06-OCT-1999;   99US-0157865P.
PR   07-OCT-1999;   99US-0158029P.
PR   08-OCT-1999;   99US-0158232P.
PR   12-OCT-1999;   99US-0158369P.
PR   13-OCT-1999;   99US-0159293P.
PR   13-OCT-1999;   99US-0159294P.
PR   13-OCT-1999;   99US-0159295P.
PR   14-OCT-1999;   99US-0159329P.
PR   14-OCT-1999;   99US-0159330P.
PR   14-OCT-1999;   99US-0159331P.
PR   14-OCT-1999;   99US-0159637P.
PR   14-OCT-1999;   99US-0159638P.
PR   18-OCT-1999;   99US-0159584P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   25-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.

  Query Match             68.4%;  Score 1017;  DB 3;  Length 352;
  Best Local Similarity   67.5%;  Pred. No. 8.4e-100;
  Matches  193;  Conservative   36;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
              :||| ::|::|| :|  |: ||| |:| |||||:||| ::|||||||| ||||  |||
Db         65 VIDGNVIAEEIRTKIISEVGKMKKAVGKVPGLAVVLVGEQRDSQTYVRNKIKACEETGIK 124

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
                |   |||| :| ::: : || | |:||||||||||| |:|:  ||| | :|||||||
Db        125 SVLAELPEDCTEGQIISVLRKFNEDTSIHGILVQLPLPQHLNESKILNMVRLEKDVDGFH 184

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
              |||:| |||:||:| || ||||||:||| | |||| || ||||||||||:| :||||:
Db        185 PLNVGNLAMRGREPLFVSCTPKGCVELLIRTGVEIAGKNAVVIGRSNIVGLPMSLLLQRH 244

Qy        183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
              :|||| ||: || || |||:|||||| |: |||||||:|||| :|||  ||:|
Db        245 DATVSTVHAFTKDPEHITRKADIVIAAAGIPNLVRGSWLKPGAVVIDVGTTPVEDSSCEF 304

Qy        243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||||||||||||  :|  |||||||||||||| ||| ||||:||||
Db        305 GYRLVGDVCYEEALGVASAITPVPGGVGPMTITMLLCNTLEAAKRI 350


RESULT 8
AAG36760
ID   AAG36760 standard; protein; 498 AA.
XX
AC   AAG36760;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 45096.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   14-MAY-1999;   99US-0134370P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
```

PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.4%; Score 1017; DB 3; Length 498;

```
  Best Local Similarity   67.5%;  Pred. No. 1.4e-99;
  Matches  193;  Conservative   36;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
              :||| ::|::|| :|  |: ||| |:| |||||:||| ::|||||||| |||| |||
Db        211 VIDGNVIAEEIRTKIISEVGKMKKAVGKVPGLAVVLVGEQRDSQTYVRNKIKACEETGIK 270

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
              |    |||| :| ::::  :  || | |:|||||||||| |:|:  ||| | :||||||||
Db        271 SVLAELPEDCTEGQIISVLRKFNEDTSIHGILVQLPLPQHLNESKILNMVRLEKDVDGFH 330

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
              |||:| |||:||:| || |||||:||| | |||| || ||||||||||:| :||||:
Db        331 PLNVGNLAMRGREPLFVSCTPKGCVELLIRTGVEIAGKNAVVIGRSNIVGLPMSLLLQRH 390

Qy        183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
              :|||| ||: || || |||:||||||| |: |||||||:|||| :|||  ||:|
Db        391 DATVSTVHAFTKDPEHITRKADIVIAAAGIPNLVRGSWLKPGAVVIDVGTTPVEDSSCEF 450

Qy        243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||||||||||||  :|  |||||||||||||| ||| ||||:||||
Db        451 GYRLVGDVCYEEALGVASAITPVPGGVGPMTITMLLCNTLEAAKRI 496


RESULT 9
AAG14693
ID   AAG14693 standard; protein; 350 AA.
XX
AC   AAG14693;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 14652.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   14-MAY-1999;   99US-0134370P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137528P.
PR   04-JUN-1999;   99US-0137502P.
PR   07-JUN-1999;   99US-0137724P.
PR   08-JUN-1999;   99US-0138094P.
PR   10-JUN-1999;   99US-0138540P.
PR   10-JUN-1999;   99US-0138847P.
PR   14-JUN-1999;   99US-0139119P.
PR   16-JUN-1999;   99US-0139452P.
PR   16-JUN-1999;   99US-0139453P.
PR   17-JUN-1999;   99US-0139492P.
PR   18-JUN-1999;   99US-0139454P.
PR   18-JUN-1999;   99US-0139455P.
PR   18-JUN-1999;   99US-0139456P.
PR   18-JUN-1999;   99US-0139457P.
PR   18-JUN-1999;   99US-0139458P.
PR   18-JUN-1999;   99US-0139459P.
PR   18-JUN-1999;   99US-0139460P.
PR   18-JUN-1999;   99US-0139461P.
PR   18-JUN-1999;   99US-0139462P.
PR   18-JUN-1999;   99US-0139463P.
PR   18-JUN-1999;   99US-0139750P.
PR   18-JUN-1999;   99US-0139763P.
PR   21-JUN-1999;   99US-0139817P.
PR   22-JUN-1999;   99US-0139899P.
PR   23-JUN-1999;   99US-0140353P.
PR   23-JUN-1999;   99US-0140354P.
PR   24-JUN-1999;   99US-0140695P.
PR   28-JUN-1999;   99US-0140823P.
PR   29-JUN-1999;   99US-0140991P.
PR   30-JUN-1999;   99US-0141287P.
PR   01-JUL-1999;   99US-0141842P.
PR   01-JUL-1999;   99US-0142154P.
PR   02-JUL-1999;   99US-0142055P.
PR   06-JUL-1999;   99US-0142390P.
PR   08-JUL-1999;   99US-0142803P.
PR   09-JUL-1999;   99US-0142920P.
PR   12-JUL-1999;   99US-0142977P.
PR   13-JUL-1999;   99US-0143542P.
PR   14-JUL-1999;   99US-0143624P.
PR   15-JUL-1999;   99US-0144005P.
PR   16-JUL-1999;   99US-0144085P.
PR   16-JUL-1999;   99US-0144086P.
PR   19-JUL-1999;   99US-0144325P.
PR   19-JUL-1999;   99US-0144331P.
PR   19-JUL-1999;   99US-0144332P.
PR   19-JUL-1999;   99US-0144333P.
PR   19-JUL-1999;   99US-0144334P.
PR   19-JUL-1999;   99US-0144335P.
PR   20-JUL-1999;   99US-0144352P.
PR   20-JUL-1999;   99US-0144632P.
PR   20-JUL-1999;   99US-0144884P.
PR   21-JUL-1999;   99US-0144814P.
PR   21-JUL-1999;   99US-0145086P.
PR   21-JUL-1999;   99US-0145088P.
PR   22-JUL-1999;   99US-0145085P.
PR   22-JUL-1999;   99US-0145087P.
PR   22-JUL-1999;   99US-0145089P.
PR   22-JUL-1999;   99US-0145192P.
PR   23-JUL-1999;   99US-0145145P.
PR   23-JUL-1999;   99US-0145218P.
PR   23-JUL-1999;   99US-0145224P.
PR   26-JUL-1999;   99US-0145276P.
PR   27-JUL-1999;   99US-0145913P.
PR   27-JUL-1999;   99US-0145918P.
```

PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.1%; Score 1012; DB 3; Length 350;
Best Local Similarity 67.1%; Pred. No. 2.9e-99;
Matches 192; Conservative 37; Mismatches 57; Indels 0; Gaps 0;

```
Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
              :||| ::|::|| :|  |: ||| |:| ||||||:||| ::|||||||||| ||||  |||
Db         63 VIDGNVIAEEIRTKIISEVGKMKKAVGKVPGLAVVLVGEQRDSQTYVRNKIKACEETGIK 122

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
              |    |||| :| :::  :  || | |:||||||||| |:::  ||| | :||||||||
Db        123 SVLAELPEDCTEGQIISVLRKFNEDTSIHGILVQLPLPQHLDESKILNMVRLEKDVDGFH 182

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
              |||:| |||:||:| || ||||||:||| | |||| || ||||||||||:| :||||:
Db        183 PLNVGNLAMRGREPLFVSCTPKGCVELLIRTGVEIAGKNAVVIGRSNIVGLPMSLLLQRH 242

Qy        183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
              :|||| ||: || || |||:|||||| |: |||||||:|||| :|||  ||:|
Db        243 DATVSTVHAFTKDPEHITRKADIVIAAAGIPNLVRGSWLKPGAVVIDVGTTPVEDSSCEF 302

Qy        243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||||||||||| :|  |||||||||||| ||| ||||:||||
Db        303 GYRLVGDVCYEEALGVASAITPVPGGVGPMTITMLLCNTLEAAKRI 348
```

RESULT 10
AAG14692
ID AAG14692 standard; protein; 352 AA.
XX
AC AAG14692;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 14651.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.

PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.1%; Score 1012; DB 3; Length 352;
Best Local Similarity 67.1%; Pred. No. 2.9e-99;
Matches 192; Conservative 37; Mismatches 57; Indels 0; Gaps 0;

```
Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
              :||| ::|::|| :|  |: ||| |:| ||||||:||| ::|||||||| |||| |||
Db         65 VIDGNVIAEEIRTKIISEVGKMKKAVGKVPGLAVVLVGEQRDSQTYVRNKIKACEETGIK 124

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
              |    |||| :| ::::  :  || | |:||||||||| |::: |||  | :||||||||
Db        125 SVLAELPEDCTEGQIISVLRKFNEDTSIHGILVQLPLPQHLDESKILNMVRLEKDVDGFH 184

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
              |||:| |||:||:| || ||||||:||| | |||| || ||||||||||:| :||||:
Db        185 PLNVGNLAMRGREPLFVSCTPKGCVELLIRTGVEIAGKNAVVIGRSNIVGLPMSLLLQRH 244

Qy        183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
              :|||| ||: || || |||:||||||| |: |||||||:|||| :||||  ||:|
Db        245 DATVSTVHAFTKDPEHITRKADIVIAAAGIPNLVRGSWLKPGAVVIDVGTTPVEDSSCEF 304

Qy        243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||||||||||||  :|  |||||||||||||| ||| ||||:||||
Db        305 GYRLVGDVCYEEALGVASAITPVPGGVGPMTITMLLCNTLEAAKRI 350
```

RESULT 11
AAG14691
ID AAG14691 standard; protein; 377 AA.
XX
AC AAG14691;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 14650.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.

PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.1%; Score 1012; DB 3; Length 377;
Best Local Similarity 67.1%; Pred. No. 3.2e-99;
Matches 192; Conservative 37; Mismatches 57; Indels 0; Gaps 0;

```
Qy          3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
              :||| ::|::|| :|  |: ||| |:| ||||||:||| ::|||||||||| |||| |||
Db         90 VIDGNVIAEEIRTKIISEVGKMKKAVGKVPGLAVVLVGEQRDSQTYVRNKIKACEETGIK 149

Qy         63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
              |    |||| :| ::::  :  || | |:||||||||| |:::  ||| | :||||||||
Db        150 SVLAELPEDCTEGQIISVLRKFNEDTSIHGILVQLPLPQHLDESKILNMVRLEKDVDGFH 209

Qy        123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
              |||:| |||:||:| || ||||||:||| | |||| || ||||||||||:| :||||:
Db        210 PLNVGNLAMRGREPLFVSCTPKGCVELLIRTGVEIAGKNAVVIGRSNIVGLPMSLLLQRH 269

Qy        183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
              :|||| ||: || || |||:||||||| |: |||||||:|||| :|||| ||:|
Db        270 DATVSTVHAFTKDPEHITRKADIVIAAAGIPNLVRGSWLKPGAVVIDVGTTPVEDSSCEF 329

Qy        243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||||||||||||  :|  |||||||||||||| ||| ||||:||||
Db        330 GYRLVGDVCYEEALGVASAITPVPGGVGPMTITMLLCNTLEAAKRI 375
```

RESULT 12
ABB60394
ID ABB60394 standard; protein; 934 AA.
XX
AC ABB60394;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7974.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL04497.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 7974; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 934 AA;

Query Match 51.3%; Score 762; DB 4; Length 934;
Best Local Similarity 55.2%; Pred. No. 7.5e-72;
Matches 160; Conservative 38; Mismatches 88; Indels 4; Gaps 3;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKM-KDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              |||| |  ||| ||||:  |:  | |     |||| :: || |:||  |:| | ||   :
Db          4 AKIISGTAVAKSIREELRNEVTAMSKQLADFVPGLRIVQVGGREDSNVYIRMKIKAATEI 63

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPL--PHHMNDENILNAVSIEKD 117
              || :  | ||   :| |:|  |    | || ||||:||:||     ::   | :||| |||
Db         64 GIDAAHVQLPRSITEVELLDKINDLNEDPRVHGIIVQMPLDCDTPIDSHRITDAVSPEKD 123

Qy        118 VDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAAL 177
              ||| | :| ||||: |    |:|||| ||:||: | |||| | ||||:||| ||| |||
Db        124 VDGLHTVNEGRLAI-GDLGGFLPCTPWGCLELIRRSGVEIAGARAVVLGRSKIVGTPAAE 182

Qy        178 LLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDD 237
              ||: |||||:: || |:  ||||| |||::  :||| :|:|||||||| :|| |||   |
Db        183 LLKWANATVTVCHSKTRNLEEITRSADILVVGIGVAEMVKGSWIKPGAVVIDCGINVKPD 242

Qy        238 PESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                | :||||| | || ::|| :|||||||||||:|||: ||: || |
Db        243 ASKASGSKLVGDVDYAEALQVAGHLTPVPGGVGPMTVAMLMKNTVRSAAR 292
```

RESULT 13
ABB48308
ID ABB48308 standard; protein; 284 AA.
XX
AC ABB48308;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #1012.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR001118.
XX
PR 11-APR-2000; 2000FR-00004629.
XX
PA (INSP) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and related
PT polypeptides.
XX
PS Claim 6; SEQ ID NO 1013; 192pp; French.
XX
CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 284 AA;

Query Match 50.8%; Score 755.5; DB 5; Length 284;
Best Local Similarity 51.6%; Pred. No. 6.2e-72;
Matches 149; Conservative 60; Mismatches 71; Indels 9; Gaps 4;

```
Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              :||||| :||:|:|:  |:|::    |  |||||:|||  : |:||||||:|  |  |:
Db          3 EIIDGKKLAKEIQEKVTREVAELVKE-GKQPGLAVVLVGDNQASRTYVRNKQKRTEEAGM 61

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
              ||  : |||: :|:::|  :   | | ::|||||||||| |:::| ::: :| :||||||
Db         62 KSVLIELPENVTEEKLLSVVEELNEDKTIHGILVQLPLPEHISEEKVIDTISFDKDVDGF 121

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              ||:|:| | : |:|  ||||||  | :||:   | :|:|||||||||||||| | | ||
Db        122 HPVNVGNLFI-GKDS-FVPCTPAGIIELIKSTGTQIEGKRAVVIGRSNIVGKPVAQLLLN 179

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              ||||:| || ||   :: ::|||:: | |:|  |:  :|||||  :||||:      :
Db        180 ENATVTIAHSRTKDLPQVAKEADILVVATGLAKFVKKDYIKPGAIVIDVGM------DRD 233

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHK 290
                :| ||| :::  : || |||||||||||||| |||:|||::||||  |
Db        234 ENNKLCGDVDFDDVVQEAGFITPVPGGVGPMTITMLLANTLKAAKRIWK 282
```

RESULT 14
ABP25547
ID ABP25547 standard; protein; 286 AA.
XX
AC ABP25547;

```
XX
DT   02-JUL-2002  (first entry)
XX
DE   Streptococcus polypeptide SEQ ID NO 270.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS   Streptococcus pyogenes.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB004789.
XX
PR   27-OCT-2000; 2000GB-00026333.
PR   24-NOV-2000; 2000GB-00028727.
PR   07-MAR-2001; 2001GB-00005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J,  Masignani V,  Margarit Y RosI,  Grandi G,  Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   N-PSDB; ABN66178.
XX
PT   New Streptococcus protein for the treatment or prevention of infection or
PT   disease caused by Streptococcus bacteria, such as meningitis, and for
PT   detecting a compound that binds to the protein.
XX
PS   Claim 1; Page 3182; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins
XX
SQ   Sequence 286 AA;

  Query Match             50.0%;  Score 743.5;  DB 5;  Length 286;
  Best Local Similarity   51.4%;  Pred. No. 1.2e-70;
  Matches  149;  Conservative   55;  Mismatches   77;  Indels    9;  Gaps    4;

Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              ::|||| :|:::::|:| ::  :|   |:||||||||||    || |||||::|   ||
Db          5 ELIDGKALAQKMQQELAAKVNNLKQKKGIVPGLAVILVGDDPASQVYVRNKERAALTVGF 64

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
              ||  | | |   ::|::  |  :|:| ::||||||||||:|:||: |: |:  :||||||
Db         65 KSETVRLSEFICQEELIAVIERYNADNTIHGILVQLPLPNHINDKKIILAIDPKKDVDGF 124

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              ||:| | |   || |  ||||| | ||||  | | ::|| ||:||||||| | | ||
Db        125 HPMNTGHL-WSGR-PLMVPCTPSGIMELLREYNVNLEGKHAVIIGRSNIVGKPMAQLLLD 182

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              ||||:: || |:: ||: | ||::| |:|  : :   :|| || :||||:|  |:
Db        183 KNATVTLTHSRTRQLEEVCRCADVLIVAIGQGHFITKQYIKDGAIVIDVGMNRDDNG--- 239

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR-IHK 290
                :|:||| ::| :::|  |||||||||||||||||  | :|| |  ||
Db        240 ---KLIGDVAFDEVAEVAAKITPVPGGVGPMTIAMLLEQTYQSALRSTHK 286


RESULT 15
ABR52676
ID   ABR52676 standard; protein; 946 AA.
XX
AC   ABR52676;
XX
DT   20-JUN-2003  (first entry)
XX
DE   Protein sequence #SEQ ID 217.
XX
KW   Multiprotein complex; eukaryote; drug target; diagnosis.
XX
OS   Saccharomyces cerevisiae.
XX
PN   EP1258494-A1.
XX
PD   20-NOV-2002.
XX
PF   20-DEC-2001; 2001EP-00130253.
XX
PR   15-MAY-2001; 2001EP-00111774.
XX
PA   (CELL-) CELLZOME AG.
XX
PI   Bauer A,  Gavin A,  Grandi P,  Krause R,  Kruse UD,  Kuester BD;
PI   Marzioch M,  Schultz JD,  Superti-Furga GD;
XX
DR   WPI; 2003-250078/25.
DR   N-PSDB; ACC60718.
XX
PT   New isolated protein complexes useful for diagnosing a disease or
PT   disorder, or as a target for an active agent of a pharmaceutical,
PT   preferably a drug target in the treatment or prevention of disease or
PT   disorder.
XX
PS   Disclosure; SEQ ID NO 217; 17pp + Sequence Listing; English.
XX
CC   The invention relates to multiprotein complexes from eukaryotes. Proteins
CC   of the invention and DNA sequences encoding them are given in records
CC   ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC   obtainable by using a protein as a bait and isolating the set of proteins
CC   which is attached thereto from cells. Such protein complexes may comprise
CC   up to 30 distinct proteins. Protein complexes of the invention are useful
CC   for diagnosing a disease or disorder, or as a target for an active agent
CC   of a pharmaceutical, preferably a drug target in the treatment or
CC   prevention of a disease or disorder. Note: The sequence data for this
CC   patent is not represented in the printed specification, but is based on
CC   sequence information supplied by the European Patent Office. The complete
CC   document is available on CD-ROM
XX
SQ   Sequence 946 AA;

  Query Match             49.8%;  Score 739.5;  DB 6;  Length 946;
  Best Local Similarity   50.5%;  Pred. No. 2e-69;
  Matches  149;  Conservative   45;  Mismatches   92;  Indels    9;  Gaps    2;

Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAI-GVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              :::|||  |:| |  || ||  ::  : |  | ||:| ||:| || |||| |:|| |  |
Db          4 QVLDGKACAQQFRSNIANEIKSIQGHVPGFAPNLAIIQVGNRPDSATYVRMKRKAAEEAG 63

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              | :  ::| | ::| |||:::   | ||  |||:|||||| |:::: | : |  ||||||
Db         64 IVANFIHLDESATEFEVLRYVDQLNEDPHTHGIIVQLPLPAHLDEDRITSRVLAEKDVDG 123

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
```

```
                 | | ||| |  :   |||:|||||| :||||:  | |:| |:||||||:||| | | ||:
Db        124 FGPTNIGELNKKNGHPFFLPCTPKGIIELLHKANVTIEGSRSVVIGRSDIVGSPVAELLK 183

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKP--------GAAIIDVGI 232
                |:||:| || |:         ||||: |:|    |:| | ||         :||||
Db        184 SLNSTVTITHSKTRDIASYLHDADIVVVAIGQPEFVKGEWFKPRDGTSSDKKTVVIDVGT 243

Qy        233 NPVDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                | | ||    |:: |||| : || |   ||||||||||||:|||: ||| :|||
Db        244 NYVADPSKKSGFKCVGDVEFNEAIKYVHLITPVPGGVGPMTVAMLMQNTLIAAKR 298

Search completed: July 24, 2004, 06:36:16
Job time : 106 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 24, 2004, 06:41:26 ; Search time 5374 Seconds
(without alignments)
2355.074 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10723061/runat_21072004_090400_7108/app_query.fasta_1.455
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10723061_@CGN_1_1_7406_@runat_21072004_090400_7108 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1486	100.0	1116	6	AX653969	AX653969 Sequence
2		1238	83.3	1083	6	AX505846	AX505846 Sequence
3		1238	83.3	1083	6	AX651582	AX651582 Sequence
4		1238	83.3	1132	8	BT000066	BT000066 Arabidops
5		1238	83.3	1498	8	AY059944	AY059944 Arabidops
6	c	1216.5	81.9	123454	8	AP005851	AP005851 Oryza sat
7	c	1116.5	75.1	91849	8	F6N23	AF058919 Arabidops
8	c	1116.5	75.1	197975	8	ATCHRIV2	AL161472 Arabidops
9	c	1054.5	71.0	92675	8	AP006402	AP006402 Lotus cor
10		1046.5	70.4	28160	8	AP006092	AP006092 Lotus cor
11		1035	69.7	1182	8	AY085538	AY085538 Arabidops
12		1034	69.6	1219	8	AF030516	AF030516 Pisum sat
13		1034	69.6	1219	8	PSA011589	AJ011589 Pisum sat
14		1030	69.3	900	8	BT004612	BT004612 Arabidops
15		1012	68.1	1118	8	AK103505	AK103505 Oryza sat
16		997.5	67.1	1282	8	BT004325	BT004325 Arabidops
17		988	66.5	1358	8	AK059603	AK059603 Oryza sat
18		949	63.9	1269	6	AX654551	AX654551 Sequence
19		892	60.0	127125	2	AC123899	AC123899 Medicago
20	c	833	56.1	107541	8	AC005499	AC005499 Arabidops
21		830.5	55.9	52211	8	AC069472	AC069472 Arabidops
22		830.5	55.9	86536	8	AP002047	AP002047 Arabidops
23		802	54.0	188080	2	AP005811	AP005811 Oryza sat
24		801	53.9	114031	8	AC139354	AC139354 Medicago
25		792	53.3	480	8	AF439725	AF439725 Zea mays
26		765.5	51.5	4160	8	AF291429	AF291429 Yarrowia
27	c	763.5	51.4	195269	6	AX417035	AX417035 Sequence
28		763.5	51.4	333050	1	AL596168	AL596168 Listeria
29		763.5	51.4	349980	6	AX417044	AX417044 Sequence
30		755.5	50.8	250050	1	AL591978	AL591978 Listeria
31		755.5	50.8	349980	6	AX641668	AX641668 Sequence
32		751.5	50.6	5635	6	AX416832	AX416832 Sequence
33	c	743.5	50.0	10029	1	AE010067	AE010067 Streptoco
34	c	743.5	50.0	10165	1	AE006583	AE006583 Streptoco
35	c	743.5	50.0	50463	1	AE014159	AE014159 Streptoco
36		743.5	50.0	310950	1	AP005143	AP005143 Streptoco
37		741	49.9	10197	1	AE009687	AE009687 Brucella
38	c	741	49.9	11059	1	AE014572	AE014572 Brucella
39		740.5	49.8	3116	5	BC045019	BC045019 Xenopus l
40		739.5	49.8	2000	6	AX594564	AX594564 Sequence
41		739.5	49.8	2000	6	AX818306	AX818306 Sequence
42		739.5	49.8	2000	6	AX829336	AX829336 Sequence
43		739.5	49.8	3529	8	SCYGR204W	Z72989 S.cerevisia
44		739.5	49.8	4883	8	YSCADE3	M12878 Saccharomyc
45		739.5	49.8	18230	8	SCPEGK620	Z49133 S.cerevisia

ALIGNMENTS

RESULT 1

```
AX653969
LOCUS       AX653969               1116 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 3839 from Patent WO03000898.
ACCESSION   AX653969
VERSION     AX653969.1  GI:29156783
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 3839 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .1116
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

Alignment Scores:
Pred. No.:              5.63e-120       Length:         1116
Score:                  1486.00         Matches:        292
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-10-723-061-18 (1-292) x AX653969 (1-1116)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 GCCAAGATCATTGATGGGAAGTTGGTGGCAAAGCAGATAAGAGAGGAAATCGCTGTTGAG 297

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 ATCGCCAAGATGAAGGATGCAATTGGGGTTGTGCCTGGGCTGGCAGTCATCCTAGTTGGG 357

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 TCAAGGAAGGATTCTCAAACGTATGTGCGCAACAAGAAGAAGGCATGCGAAGCGGTTGGT 417

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        418 ATCAAGTCATATGAGGTTAATTTGCCGGAAGACAGCTCTGAGGATGAGGTTCTCAAGCAC 477

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        478 ATCGCAACATTTAACAGTGATCCGTCGGTGCATGGCATCTTGGTTCAGTTGCCCCTACCT 537

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        538 CATCATATGAATGATGAGAACATTTTGAATGCTGTTAGTATTGAGAAGGATGTTGATGGA 597

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 TTTCATCCACTGAACATTGGACGACTTGCAATGCAAGGTCGGGATCCGTTCTTTGTTCCA 657

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 TGCACCCCTAAAGGATGCATGGAATTACTACACAGATATGGAGTTGAAATCAAAGGGAAG 717

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 AGAGCTGTTGTAATTGGGCGGAGCAATATTGTGGGGATGCCTGCTGCATTATTACTGCAA 777

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 AAAGCCAACGCAACTGTTAGCATTGTACATTCAAATACCAAGAAGCCTGAGGAAATAACA 837
```

```
Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 AGACAAGCAGATATTGTTATCGCAGCTGTTGGAGTTGCTAATCTGGTCAGAGGGAGTTGG 897

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 ATAAAGCCTGGTGCTGCTATTATTGATGTTGGCATCAATCCAGTTGATGATCCAGAAAGC 957

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 CCTCGAGGTTATCGGCTGGTTGGAGATGTGTGCTACGAGGAGGCCTCCAAGATTGCAGGA 1017

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 CTAATCACGCCAGTTCCTGGTGGTGTCGGGCCAATGACAATTGCGATGCTTTTGTCGAAC 1077

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||||||||||||||||||||||||||||||||
Db       1078 ACACTTGAGTCAGCTAAAAGGATACACAAATTCAAA 1113
```

```
RESULT 2
AX505846
LOCUS       AX505846               1083 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 541 from Patent WO0216655.
ACCESSION   AX505846
VERSION     AX505846.1  GI:23387083
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
  TITLE     Stress-regulated genes of plants, transgenic plants containing
            same, and methods of use
  JOURNAL   Patent: WO 0216655-A 541 28-FEB-2002;
            The Scripps Research Institute (US) ; Syngenta Participations AG
            (CH)
FEATURES             Location/Qualifiers
     source          1. .1083
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:              2.01e-98        Length:         1083
Score:                  1238.00         Matches:        231
Percent Similarity:     91.44%          Conservative:   36
Best Local Similarity:  79.11%          Mismatches:     25
Query Match:            83.31%          Indels:         0
DB:                     6               Gaps:           0

US-10-723-061-18 (1-292) x AX505846 (1-1083)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||  :::|||||||||||||  |||||||||:::|||||:::||||||   :::||||
Db        205 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 264

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::|||||:::::|||||||||:::|||||||||||||||||||||||||||||
Db        265 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 324

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              |||||||||||  |||||||||||||||||||||||||||||:::::||||||||||||
Db        325 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 384

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              ||||||||||:::|||||||  |||  |||||||||||||||:::||||||||||||
Db        385 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 444
```

```
Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db        445 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 504

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db        505 TCGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATAGAGAAAGATGTTGACGGA 564

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
Db        565 TTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGAGAACCCTTATTCGTTCCT 624

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db        625 TGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAACATTGAAATCAAAGGAAAG 684

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
Db        685 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCAGCTGCTCTTTTACTGCAG 744

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db        745 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 804

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db        805 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 864

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db        865 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGAGGATCCAAGTGCT 924

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db        925 GCGCGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCA 984

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db        985 GCCATCACACCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1044

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
Db       1045 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1080

RESULT 3
AX651582
LOCUS       AX651582                1083 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 393 from Patent WO03000898.
ACCESSION   AX651582
VERSION     AX651582.1  GI:29154400
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 393 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .1083
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:              2.01e-98        Length:         1083
Score:                  1238.00         Matches:        231
Percent Similarity:     91.44%          Conservative:   36

Best Local Similarity:  79.11%          Mismatches:     25
Query Match:            83.31%          Indels:         0
DB:                     6               Gaps:           0

US-10-723-061-18 (1-292) x AX651582 (1-1083)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db        205 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 264

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db        265 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 324

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db        325 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 384

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db        385 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 444

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db        445 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 504

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db        505 TCGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATAGAGAAAGATGTTGACGGA 564

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
Db        565 TTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGAGAACCCTTATTCGTTCCT 624

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db        625 TGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAACATTGAAATCAAAGGAAAG 684

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
Db        685 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCAGCTGCTCTTTTACTGCAG 744

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db        745 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 804

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db        805 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 864

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db        865 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGAGGATCCAAGTGCT 924

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db        925 GCGCGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCA 984

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db        985 GCCATCACACCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1044

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
Db       1045 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1080

RESULT 4
BT000066
LOCUS       BT000066                1132 bp    mRNA    linear   PLN 19-SEP-2002
DEFINITION  Arabidopsis thaliana Unknown protein (At4g00620) mRNA, complete
            cds.
ACCESSION   BT000066
VERSION     BT000066.1  GI:23197715
```

```
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1132)
  AUTHORS   Nguyen,M., Karlin-Neumann,G., Southwick,A., Tripp,M., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2002) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     e-mail for correspondence: arab@sequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis  Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pENTR (ORF) clones using the RAFL cDNAs: Nguyen,M.,
            Southwick,A., Tripp,M., Palm,C.J., Jones,T., Wu,T., Chen,H.,
            Cheuk,R., Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W.,
            Lee,J.M., Kim,C.J., Quach,H.L., Shinn,P., Tang,C.C., Toroumi,M.,
            Wallender,E.K., Wong,C., Wu,H.C., Yamada,K., Yu,G., Yuan,S.,
            Ecker,J., Theologis,A. and Davis,R.W.

            Nguyen,M., (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
            equally to this work.  Shinozaki,K. (RIKEN GSC) and Davis,R.W.
            (SSP/Stanford) contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1132
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /clone="U16952"
                     /ecotype="Columbia"
                     /note="This clone is in pENTR/SD-dTopo   This is a cloned
                     PCR product using RIKEN clone RAFL06-08-M11 (AY059944) as
                     a template"
     gene            1. .1132
                     /gene="At4g00620"
                     /note="synonym: F6N23.26"
     CDS             1. .1083
                     /gene="At4g00620"
                     /codon_start=1
                     /product="Unknown protein"
                     /protein_id="AAN15385.1"
                     /db_xref="GI:23197716"
                     /translation="MASMMFTDCSSTTTSRLIHLNRSSGTFLLRQCVGQLRLQTTASG
                     RGCCIRSSSSPISSISADTKSEGGAIVIDGKAVAKKIRDEITIEVSRMKESIGVIPGL
                     AVILVGDRKDSATYVRNKKKACDSVGIKSFEVRLAEDSSEEEVLKSVSGFNDDPSVHG
                     ILVQLPLPSHMDEQNILNAVSIEKDVDGFHPLNIGRLAMRGREPLFVPCTPKGCIELL
                     HRYNIEIKGKRAVVIGRSNIVGMPAALLLQREDATVSIIHSRTKNPEEITREADIIIS
                     AVGQPNMVRGSWIKPGAVLIDVGINPVEDPSAARGYRLVGDICYEEASKVASAITPVP
                     GGVGPMTIAMLLSNTLTSAKRIHNFQ"
ORIGIN

Alignment Scores:
Pred. No.:              2.12e-98        Length:         1132
Score:                  1238.00         Matches:        231
Percent Similarity:     91.44%          Conservative:   36
Best Local Similarity:  79.11%          Mismatches:     25
Query Match:            83.31%          Indels:         0
DB:                     8               Gaps:           0

US-10-723-061-18 (1-292) x BT000066 (1-1132)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::|||||||||||||   |||||||||:::||||||:::||||||   :::|||
Db        205 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 264

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::|||||:::::::||||||||:::||||||||||||||||||||||||||||
Db        265 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 324

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
                |||||||||||   |||||||||||||||||||||||||||||:::::||||||
Db        325 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 384

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              ||||||||||:::||||||   |||   ||||||||||||||:::||||||||||||
Db        385 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 444

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::   ||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        445 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 504

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              ||||||:::::::::|||||||||||||||||||||||||||||||||||||||||||||
Db        505 TCGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATAGAGAAAGATGTTGACGGA 564

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||||||||||||||:::||||||:::|||   |||||||||
Db        565 TTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGAGAACCCTTATTCGTTCCT 624

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||:::|||||||||||||||||   :::|||||||||||||||||
Db        625 TGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAACATTGAAATCAAAGGAAAG 684

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCAGCTGCTCTTTTACTGCAG 744

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::|||||||||||||||:::||||||   ||||||   ||||||||||||||||
Db        745 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 804

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::||||||||||:::|||:::||||||||   |||:::|||||||||||||||||
Db        805 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 864

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||   :::|||||||||||||||||||||||||::||||||   :::
Db        865 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGAGGATCCAAGTGCT 924

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              |||||||||||||||||||||||:::|||||||||||||||||||:::|||
Db        925 GCGCGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCA 984

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 GCCATCACACCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1044

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||   ||||||||||||||||||   |||:::
Db       1045 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1080

RESULT 5
AY059944
LOCUS       AY059944               1498 bp    mRNA    linear   PLN 04-NOV-2001
DEFINITION  Arabidopsis thaliana Unknown protein (At4g00620; F6N23.26) mRNA,
            complete cds.
ACCESSION   AY059944
VERSION     AY059944.1  GI:16649148
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1498)
AUTHORS Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K., Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2001) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M, (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES Location/Qualifiers
source 1. .1498
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="IV"
/clone="RAFL06-08-M11"
/note="This clone is in pBluescript"
gene 1. .1498
/gene="At4g00620; F6N23.26"
CDS 187. .1269
/gene="At4g00620; F6N23.26"
/codon_start=1
/product="Unknown protein"
/protein_id="AAL24426.1"
/db_xref="GI:16649149"
/translation="MASMMFTDCSSTTTSRLIHLNRSSGTFLLRQCVGQLRLQTTASG RGCCIRSSSSPISSISADTKSEGGAIVIDGKAVAKKIRDEITIEVSRMKESIGVIPGL AVILVGDRKDSATYVRNKKKACDSVGIKSFEVRLAEDSSEEEVLKSVSGFNDDPSVHG ILVQLPLPSHMDEQNILNAVSIEKDVDGFHPLNIGRLAMRGREPLFVPCTPKGCIELL HRYNIEIKGKRAVVIGRSNIVGMPAALLLQREDATVSIIHSRTKNPEEITREADIIIS AVGQPNMVRGSWIKPGAVLIDVGINPVEDPSAARGYRLVGDICYEEASKVASAITPVP GGVGPMTIAMLLSNTLTSAKRIHNFQ"

ORIGIN

Alignment Scores:

Pred. No.:	2.97e-98	Length:	1498
Score:	1238.00	Matches:	231
Percent Similarity:	91.44%	Conservative:	36
Best Local Similarity:	79.11%	Mismatches:	25
Query Match:	83.31%	Indels:	0
DB:	8	Gaps:	0

US-10-723-061-18 (1-292) x AY059944 (1-1498)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::|||||||||||||   |||||||||:::||||||:::||||||   :::|||
Db        391 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 450

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::||||||::::::||||||||::||||||||||||||||||||||||||||
Db        451 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 510

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||   ||||||||||||||||||||||||||||||:::::|||||
Db        511 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 570

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::||||||   |||   ||||||||||||||:::||||||||||||
Db        571 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 630

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::   ||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        631 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 690

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||||:::::::::||||||||||||||||||||||||||||||||||||||||||||
Db        691 TCGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATAGAGAAAGATGTTGACGGA 750

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||||||||||||||:::|||||:::|||   |||||||||
Db        751 TTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGAGAACCCTTATTCGTTCCT 810

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||:::||||||||||||||||   :::|||||||||||||||||
Db        811 TGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAACATTGAAATCAAAGGAAAG 870

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        871 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCAGCTGCTCTTTTACTGCAG 930

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::|||||||||||||:::|||||||   ||||||   ||||||||||||||||
Db        931 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 990

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||:::|||:::|||||||||       |||:::|||||||||||||||
Db        991 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 1050

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||   :::|||||||||||||||||||||||:::|||||   :::
Db       1051 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGAGGATCCAAGTGCT 1110

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||||:::|||||||||||||||||||:::||||
Db       1111 GCGCGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCA 1170

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1171 GCCATCACACCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1230

Qy        281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||||||   ||||||||||||||||||   |||:::
Db       1231 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1266
```

RESULT 6
AP005851/c
LOCUS AP005851 123454 bp DNA linear PLN 01-AUG-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2, BAC clone:OSJNBb0088N06, complete sequence.
ACCESSION AP005851
VERSION AP005851.2 GI:33383227
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC clone:OSJNBb0088N06
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 123454)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Jul 31, 2003 this sequence version replaced gi:24371211.
            The orientation of the sequence is from M13rev to -21M13 of the BAC
            clone.
FEATURES             Location/Qualifiers
     source          1. .123454
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="2"
                     /clone="OSJNBb0088N06"
ORIGIN

Alignment Scores:
Pred. No.:                4.48e-94      Length:          123454
Score:                    1216.50       Matches:         292
Percent Similarity:       37.87%        Conservative:    0
Best Local Similarity:    37.87%        Mismatches:      0
Query Match:              81.86%        Indels:          479
DB:                       8             Gaps:            3

US-10-723-061-18 (1-292) x AP005851 (1-123454)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77773 GCCAAGATCATTGATGGGAAGTTGGTGGCAAAGCAGATAAGAGAGGAAATCGCTGTTGAG 77714

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77713 ATCGCCAAGATGAAGGATGCAATTGGGGTTGTGCCTGGGCTGGCAGTCATCCTAGTTGGG 77654

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77653 TCAAGGAAGGATTCTCAAACGTATGTGCGCAACAAGAAGAAGGCATGCGAAGCGGTTGGT 77594

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77593 ATCAAGTCATATGAGGTTAATTTGCCGGAAGACAGCTCTGAGGATGAGGTTCTCAAGCAC 77534

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77533 ATCGCAACATTTAACAGTGATCCGTCGGTGCATGGCATCTTGGTTCAGTTGCCCCTACCT 77474

Qy        101 His---------------------------------------------------------- 101
              |||
Db      77473 CATGTAAGTTTTCAACAATTTTCTTTGGGCCAATGTCTCTGTATAACCTATCCACTGATA 77414

Qy        101 ---------------------------------------------------------- 101

Db      77413 TTGAATTTTGGCCCTTACCAACTGAACTGGTAGTTAAATGTATTTTTAGTGAGGTACATT 77354

Qy        101 ---------------------------------------------------------- 101

Db      77353 CCTAGATATGTATGGCAGTATGGCACACCATGGAAACATTACTTTATGCACAATATTTTC 77294

Qy        101 ---------------------------------------------------------- 101

Db      77293 TGGAAATATTGTGATATTGGTGCCATCTCCAATTTGTACACTAAATGTTTCTATTTATGT 77234

Qy        101 ---------------------------------------------------------- 101

Db      77233 TAACTGGTTTGATGCTGCTCTTGTGAAAGATATCTGGTCTTAGAAAAATGTGTGCAAACA 77174

Qy        101 ---------------------------------------------------------- 101

Db      77173 GATCCTTTGTTTTAGGTAGTCCGTCGCATTCCCTTTGTGGTCTTTCGAGTTTGTTGTCAG 77114

Qy        101 ---------------------------------------------------------- 101

Db      77113 CATTTGTTTTTAATTAGATGGTCTATACTGGTGCTTGCAGTTCTCTGTTAATCTTGTTGC 77054

Qy        101 ---------------------------------------------------------- 101

Db      77053 ATTGTTTGTAGCATATTAATGATGGGGATATTTTTGAATAATTTTTTGATCAAGATACTT 76994

Qy        101 ---------------------------------------------------------- 101

Db      76993 TGAATGATTTTACCTGTTAGCATTAAAAGAGGTATTTGCCTGCTAGCATCACACAAGTTT 76934

Qy        102 ----------------------------------------------------------HisMet 103
                                                                         ||||||
Db      76933 TCCATGTCTTCCATCCACTGAGCATCTGTATTTCTGCTCTCTGATGGTTTGCAGCATATG 76874

Qy        104 AsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHisPro 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76873 AATGATGAGAACATTTTGAATGCTGTTAGTATTGAGAAGGATGTTGATGGATTTCATCCA 76814

Qy        124 LeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThrPro 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76813 CTGAACATTGGACGACTTGCAATGCAAGGTCGGGATCCGTTCTTTGTTCCATGCACCCCT 76754

Qy        144 LysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAlaVal 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76753 AAAGGATGCATGGAATTACTACACAGATATGGAGTTGAAATCAAAGGGAAGAGAGCTGTT 76694

Qy        164 ValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln--------- 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76693 GTAATTGGGCGGAGCAATATTGTGGGGATGCCTGCTGCATTATTACTGCAAGTATGAATT 76634

Qy        180 ---------------------------------------------------------- 180

Db      76633 AATTCTCACTTCAATTAGTTGGAGAAATTTTATTGGAGTGTTTAATTAACATGTTTTTAT 76574

Qy        181 ------------------LysAlaAsnAlaThrValSerIleValHisSerAsnThrLys 194
                               ||||||||||||||||||||||||||||||||||||||||||
Db      76573 GCCATATGTGATTACTAGAAAGCCAACGCAACTGTTAGCATTGTACATTCAAATACCAAG 76514

Qy        195 LysProGluGluIleThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsn 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76513 AAGCCTGAGGAAATAACAAGACAAGCAGATATTGTTATCGCAGCTGTTGGAGTTGCTAAT 76454

Qy        215 LeuValArgGlySerTrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnPro 234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76453 CTGGTCAGAGGGAGTTGGATAAAGCCTGGTGCTGCTATTATTGATGTTGGCATCAATCCA 76394

Qy        235 ValAsp------------------------------------------------------ 236
              ||||||
Db      76393 GTTGATGTAAGACACCTACTTTTCTGATTGGCTTCCTTAATTCTTCCTCCCTAACATTCA 76334

Qy        236 ---------------------------------------------------------- 236

Db      76333 ACAGATCTGTTTGCTTTTAATGCCAACCCATTTCTAGATGGATAAGTTTTATGCATGAAA 76274

Qy        236 ---------------------------------------------------------- 236

Db      76273 ACTGGGCATGTCATAGCCAAAGTCCCTTATATTTTGGGACGGAGGTAGTTTGAACCTGCT 76214

Qy        236 ---------------------------------------------------------- 236

Db      76213 AATAAGATTTCCAATGTTACTGGTTCCCTGGAATATCACCACATTGGTGGGATCCAGATG 76154

Qy        236 ---------------------------------------------------------- 236

Db      76153 GCTAATGGTGTATTTGTGTTGTTAATAAAATAATATATAATGCTTTTGCACTTCATGTTT 76094

Qy        236 ---------------------------------------------------------- 236

Db      76093 CTAATCATGTTTGGTTCATATATGCAAAATGTCACATTCTTAGAAGGAATATAGTTTTTT 76034

Qy        236 ---------------------------------------------------------- 236
```

```
Db      76033 TTTTCAAGAACAAAGGGAACATTATTGTCTAGATCCTCGCAGCTGGGATCGAAGGCAGGT 75974

Qy        236 ------------------------------------------------------------ 236

Db      75973 AAGGAACTACACCAGTAGACCCTCCTGTGATCCCACTGTTGGACACCCATTTATAAAAAC 75914

Qy        236 ------------------------------------------------------------ 236

Db      75913 TTAAATTGCCTTATTAAATTTTGATGATCCCATATCTTTGCTGGTAAATTTCTTCACATT 75854

Qy        236 ------------------------------------------------------------ 236

Db      75853 GTATATGTGTTTGGGCAATATGTGATGATTAATCAATTTAGTACAAGCCTTCACTTAAAT 75794

Qy        236 ------------------------------------------------------------ 236

Db      75793 GTTGGCTTGTGCATATCTTTAAGTTTTGTTGTATGTTTTTGTATCCTAACTCAGTTAATA 75734

Qy        236 ------------------------------------------------------------ 236

Db      75733 CCTCAGAAGGTTCCAAAGAAATTGGTGACTTGGCTGCTTCTTTCCCTACACCCCCTTCCT 75674

Qy        237 ----------------------------------------------AspProGluSerPro 241
                                                             ||||||||||||||||
Db      75673 GATTTCCTTGCATGCTATGCTTACTAGTGATTTGAAATGTCTCAGGATCCAGAAAGCCCT 75614

Qy        242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75613 CGAGGTTATCGGCTGGTTGGAGATGTGTGCTACGAGGAGGCCTCCAAGATTGCAGGACTA 75554

Qy        262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75553 ATCACGCCAGTTCCTGGTGGTGTCGGGCCAATGACAATTGCGATGCTTTTGTCGAACACA 75494

Qy        282 LeuGluSerAlaLysArgIleHisLysPheLys 292
               |||||||||||||||||||||||||||||||
Db      75493 CTTGAGTCAGCTAAAAGGATACACAAATTCAAA 75461
```

RESULT 7
F6N23/c
LOCUS F6N23 91849 bp DNA linear PLN 12-NOV-1999
DEFINITION Arabidopsis thaliana BAC F6N23.
ACCESSION AF058919
VERSION AF058919.2 GI:6382042
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 91849)
AUTHORS Geisel,C.
TITLE The sequence of A. thaliana F6N23
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 91849)
AUTHORS Washington University Genome Sequencing Center.
TITLE The A. thaliana Genome Sequencing Project
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 91849)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 91849)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Nov 12, 1999 this sequence version replaced gi:3047100.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: rwilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEFHBORING COSMID INFORMATION:

The 5' clone is F5I10, 200 bp overlap;3' clone is F15P23, 900 bp overlap. Actual start of this clone is at base position 104960 of CELF5I10; actual end is at 91149 of CELF6N23

The clone sequenced to the left is F5I10. The actual start of this clone is at base position 104960 of F5I10; actual end is at base position 91149 of F6N23.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES Location/Qualifiers
source 1. .91849
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="IV"
/map="unknown"
/clone="F6N23"
gene complement(187. .907)
/gene="F6N23.23"
CDS complement(join(187. .347,396. .512,598. .907))
/gene="F6N23.23"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC13624.2"
/db_xref="GI:6382044"
/translation="MENLSNPDENDDHQSPRSIDQNDQSAVETPVYSTMSIDSFVYPR
TCSETTSGFSDQIDETNSFCSEASPCNWPVLTESKSSKCLSGLEMQSNECLVVQEISE
PELETMKERFAKLLLGEDMSGSGKGVCTAVTISNAITNLYDNCKNDTSFDNVCSYSVW
TESEVRAVRNREESIVEERNELSFICMRLHCRIHP"
gene complement(4919. .7497)
/gene="F6N23.22"
CDS complement(join(4919. .4993,5078. .5488,5567. .6055,
6157. .6213,6498. .6594,6935. .7222,7349. .7497))
/gene="F6N23.22"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC13623.1"
/db_xref="GI:3047112"
/translation="MSLTMADGVEAAAGRSKRQNSLLRKQLALAVRSVQWSYAIFWSS
SLTQPGVLEWGEGCYNGDMKKRKKSYESHYKYGLQKSKELRKLYLSMLEGDSGTTVST
THDNLNDDDDNCHSTSMMLSPDDLSDEEWYYLVSMSYVFSPSQCLPGRASATGETIWL
CNAQYAENKLFSRSLLARTVVCFPYLGGVIELGVTELISEDHNLLRNIKSCLMEISAH
QDNDDEKKMEIKISEEKHQLPLGISDEDLHYKRTISTVLNYSADRSGKNDKNIRHRQP
NIVTSEPGSSFLRWKQCEQQVSGFVQKKKSQNVLRKILHDVPLMHTKRMFPSQNSGLN
QDDPSDRRKENEKFSVLRTMVPTVNEVDKESILNNTIKYLQELEARVEELESCMGSVN
```
```

```
                FVERQRKTTENLNDSVLIEETSGNYDDSTKIDDNSGETEQVTVFRDKTHLRVKLKETE
                VVIEVRCSYRDYIVADIMETLSNLHMDAFSVRSHTLNKFLTLNLKAKFRGAAVASVGM
                IKRELRRVIGDLF"
gene            10235. .12675
                /gene="F6N23.1"
CDS             join(10235. .10495,10693. .10896,10998. .11192,11341. .11505,
                11633. .11893,11998. .12216,12315. .12534,12617. .12675)
                /gene="F6N23.1"
                /note="similar to the family of glycosyl hydrolases"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13634.1"
                /db_xref="GI:3047123"
                /translation="MAIRLNHSVIPVSVKLGAPTRVSARSSLPFSVGDWRGVSTFSGA
                RPLVLAKVKLRAESTEEDRVPIDDDDDSTDQLVSVEFHLDDSLLGVIDMNSEVVEPEE
                LLDQLRTLKSVNVDGVMVDCWWGIVESHTPQVYNWSGYKKLFQMIRELGLKIQVVMSF
                HECGGNVGDDVHIQIPEWVREIGQSNPDIYFTDSAGRRNTECLTWGIDKQRVLRGRTA
                LEVYFDYMRSFRVEFDEFFEEKIIPEIEVGLGPCGELRYPSYPAQFGWKYPGIGEFQC
                YDKYLMNSLKEAAEVRGHSFWGRGPDNTETYNSTPHGTGFFRDGGDYDSYYGRFFLNW
                YSRVLIDHGDRVLAMANLAFEGTCIAAKLSGIHWWYKTASHAAELTAGFYNSSNRDGY
                GPIAAMFKKHDAALNFTCVELRTLDQHEDFPEALADPEGLVWQVLNAAWDASIPVASE
                NALPCYDREGYNKILENAKPLTDPDGRHLSCFTYLRLNPTLMESQNFKEFERFLKRMH
                GEAVPDLGLAPGTQETNPE"
gene            complement(13027. .14791)
                /gene="F6N23.21"
CDS             complement(join(13027. .13335,13480. .13615,13712. .13840,
                13943. .14001,14087. .14791))
                /gene="F6N23.21"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13622.1"
                /db_xref="GI:3047111"
                /translation="MSILCCVPVLECVYCLGCTHWLWKKCLYSAGHESENWGLATSDE
                FEPIPRICRLILAVYEENLHDPMWAPPDGYGIDPNHVILKKDYDQTEGRVTPYMIYLD
                HENGDVVLAIRGLNLAKECDYAVLLDNKLGQTKFDGGYVHNGLLKAAMWVFEEEHVVL
                RELLEANPSYSLTFVGHSLGAGVVSLLVLFVIQNRVRLGNIERKRIRCFAIAPPRCMS
                LHLAVTYADVINSVVLQDDFLPRTTTALENVFKSIICLPCLLCLTCLKDTFTFEERKL
                KDARRLYAPGRLYHIVVRKPLRLGRYPPVVRTAVPVDGRFEQIVLSCNATADHAIIWI
                ERESQRALDKIVRQKSIVEDHDEEYRAAIMKAASLNIPMSPSPSYGTFHDTEEGESSA
                GSGMEGSPSGWSFKGMRRKWDQFIDCHFPVNDNSEHMIFKNQESQALLCGQET"
gene            15539. .16687
                /gene="F6N23.2"
CDS             join(15539. .15554,15672. .15764,15953. .16026,16133. .16208,
                16307. .16353,16466. .16687)
                /gene="F6N23.2"
                /note="Similar to protein kinase; coded for by A. thaliana
                cDNA H36947; coded for by A. thaliana cDNA H37158"
                /codon_start=1
                /protein_id="AAC13612.1"
                /db_xref="GI:3047101"
                /translation="MNTESVVEFLGNVTLLQRLPSSSLKRISEVVVFKGYDRGDYVVR
                ENQNVDGVYFLLQGQVRAQVLRSAGEENYQEFPLKRYDFFGHGIFGDVYSADVVAVTE
                LTCLLLMSDHRALLEIKSVSDSDKERCLVEDILYLEPLDVRVSPNCPSFSSISVLMIK
                VLTIHFYHHSLESEL"
gene            16976. .19151
                /gene="F6N23.3"
CDS             join(16976. .17111,17262. .17415,17785. .17889,18316. .18417,
                18618. .18666,19056. .19151)
                /gene="F6N23.3"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13635.1"
                /db_xref="GI:3047124"
                /translation="MEIYSNHCDTTVYASQALAAASKTVETMKIVHNFHCYFLLVGDI
                NIPIIYDVNRLRDGNNFATRSVDARQKGKTIFTLFASFQVSCYIPEIFHYESLDYRNQ
                VATEISVPFPIDIRFCEPNRSTKQNKSPPRCVVAFASDLIFATISLNPHRREGMSVAA
                LSLDHSMWFHRPVRADDWLLFVIVSPTATESRGFATGKMFNRKGEVRCCIEEE"
gene            complement(20843. .21366)
                /gene="F6N23.20"
CDS             complement(join(20843. .20944,21002. .21056,21103. .21155,
                21292. .21366))
                /gene="F6N23.20"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13621.1"
                /db_xref="GI:3047110"
                /translation="MVATAVASYFLLTGDYGPEPNVLDPYLSSSMILLLVEVLVIRMF
                SYWWVMLHLFMLVIEHWIKRKILAAQDSVKQFIFGPKGEPSGKEPSDTAK"
gene            complement(22410. .24753)
                /gene="F6N23.19"
CDS             complement(join(22410. .22521,23220. .23538,23620. .23939,
                24009. .24137,24240. .24352,24439. .24548,24639. .24753))
                /gene="F6N23.19"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13637.1"
                /db_xref="GI:3047126"
                /translation="MSCPSYINVSLKPFFHLMQDQILTNVVKKYQGRNWKRIAECLPG
                SEENRRNDVQCQHRWLKVLDPSLQKGAWKKEEDELLSELVKDYMENDRPPWSKISKEL
                PGRIGKQCRERWHNHLNPTIIKSPWTREEELILVQAQRGNGNKWAEIAKLLPGRTENN
                IKNHWNCSVKKRLEQFPSNLFSGVVYGSKPSSGFEYNFFNQRNTMVESCITSQIKEAA
                KSPQRDFLDLTLGLNWRSISSSTSSLRGEESVSSSVDSVCARLNACLETPQNSNNDTV
                CVKEVREMKERLRMAARTFDTPSIISKTSSPASGLKRLRQKYDTPFPTDARSHMSSEE
                DHSVSASPSSKYRFVKRNTCSGSKPLERRLDFDFLLWDEHGRRNGIVNFSVRILPQKS
                DLKSGLVRPFWLR"

Alignment Scores:
Pred. No.:              1.56e-85        Length:         91849
Score:                  1116.50         Matches:        231
Percent Similarity:     68.38%          Conservative:   35
Best Local Similarity:  59.38%          Mismatches:     26
Query Match:            75.13%          Indels:         99
DB:                     8               Gaps:           3

US-10-723-061-18 (1-292) x F6N23 (1-91849)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::|||||||||||||   |||||||||:::||||||:::|||||||   :::|||
Db      48245 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAGATTAGAGATGAAATCACAATTGAA 48186

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::|||||:::::::|||||||||:::|||||||||||||||||||||||||||||
Db      48185 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 48126

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||   |||||||||||||||||||||||||||||||:::::|||||||
Db      48125 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 48066

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::||||||   |||   |||||||||||||||:::||||||||||||
Db      48065 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 48006

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::  ||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db      48005 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 47946

Qy        101 His------------------------------------------------------------ 101

Db      47945 TC-GGTAAACAAAGTCTAATAACTTTCTTCGGAACTTAAATAAACAGTGAATTACAATAG 47887

Qy        102 --------------------HisMetAsnAspGluAsnIleLeuAsnAlaValSerIle 114
                                  ||||||:::::::::|||||||||||||||||||||||||
Db      47886 TGTTTTGGGATGTTGTTGCAGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATA 47827

Qy        115 GluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArg 134
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db      47826 GAGAAAGATGTTGACGGATTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGA 47767

Qy        135 AspProPhePheValProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGly 154
              :::|||   ||||||||||||||||||||||||||:::|||||||||||||||||||
Db      47766 GAACCCTTATTCGTTCCTTGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAAC 47707

Qy        155 ValGluIleLysGlyLysArgAlaValValIleGlyArgSerAsnIleValGlyMetPro 174
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47706 ATTGAAATCAAAGGAAAGAGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCA 47647
```

```
Qy        175 AlaAlaLeuLeuLeuGln---------------------------------------- 180
              ||||||||||||||||||
Db      47646 GCTGCTCTTTTACTGCAGGTAATGATGATGATACCTTCTTTAAGATGCCGTTTTTTCTGG 47587

Qy        180 ------------------------------------------------------------ 180

Db      47586 TTGTGACATTCGAAAGTTAGTTACATGTTTGGTCTTTATTGTTGTGTTTTGTATTTGTAG 47527

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::||||||||||||||||:::||||||   ||||||   ||||||||||||||
Db      47526 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 47467

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::||||||||||:::|||:::|||||||||       |||:::|||||||||||||||
Db      47466 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 47407

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAsp------------ 236
              ||||||||||||||   :::||||||||||||||||||||||||
Db      47406 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGA-GGTGAGTGGAAT 47348

Qy        236 ------------------------------------------------------------ 236

Db      47347 TTTCTTTTTGACATATGGTCATGGCGTAATTCAGTGATATAATAAGTTTCAAATCAACCA 47288

Qy        237 ---------------------------------------AspProGluSerProArgGly 243
                                                      ||||||   :::   ||||||
Db      47287 GTATGCAACATTCATACATGTACATAATATGAATTGCAGGATCCAAGTGCTGCGCGTGGA 47228

Qy        244 TyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIleThr 263
              ||||||||||||||||:::|||||||||||||||||||:::|||      ||||||
Db      47227 TATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCAGCCATCACA 47168

Qy        264 ProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeuGlu 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47167 CCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAACACTTTAACA 47108

Qy        284 SerAlaLysArgIleHisLysPheLys 292
              ||||||||||||||||||   |||:::
Db      47107 TCAGCTAAGAGGATTCACAACTTCCAG 47081
```

```
RESULT 8
ATCHRIV2/c
LOCUS       ATCHRIV2              197975 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 2.
ACCESSION   AL161472
VERSION     AL161472.2  GI:7267383
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 197975)
  AUTHORS   Lamar,B., Stoneking,T., Stumpf,J., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197975)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV1 at the 5' end and an
            overlap with ATCHRIV3 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .197975
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            10352. .17547
                     /gene="AT4g00450"
     CDS             join(10352. .10384,10432. .10545,10605. .10709,10849. .11018,
                     11079. .11423,11475. .13674,13735. .13823,13921. .16530,
                     16677. .16862,16976. .17042,17092. .17547)
                     /gene="AT4g00450"
                     /note="Contains Aminoacyl-transfer RNA synthetases
                     class-II signatures AA1343-1352;Prokaryotic membrane
                     lipoprotein lipid attachment site AA349-359
                     contains EST gb:AI999580.1"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB80854.1"
                     /db_xref="GI:7267384"
                     /db_xref="SPTREMBL:O23057"
                     /translation="MLSPLFVISIGAVRKCLRAINESRALKRKVIKRLSDSNPFQFIF
                     CAEFRAGQVYGVPLSGSLLCKPGFPEQRSCGEETKKRWIESLSQQHKRLRSLADNIPG
                     YRRKTLFEVLIRNNVPLLRATWFIKVTYLNQVHCWAINWCDLVLLLFLQEHLTRHKLL
                     GVSNGQKMLLNICNTSWMNFCHGIAHFLLSKLEIGHHRCFIQDQCKRIVQHQQAFTAR
                     KHLYILNGGIWCVFYSGTMLKGFFFLISLLIGFSSSYSFVNFHFALLQEKEIFEILQL
                     LLPIVYGVLESIVLSQTYVQSLVAIAVRFIQEPAPGGSDLVDNSRRAYTLSALIEMVR
                     YLVLAAPDTFVASDFFPLPPSVAACGPNDVSYTSKAYENLEKLRSNSAEISAQFQGRG
                     VLSRFEFLSFDYTISTIQRSADDLAKIASAGYPQHNVAKAVQALDKALSDGDIRAAYS
                     YLFEDLCNGAVDEAWITDVSPCLRSSLRWIGAISTSFVCSVFFLIEWATCDFRDFRAG
                     VPKDIKFSGRKDCSQVYLVIQLLKQKILGGEFTARKGKNCRNNFLGVSKPSGSMDAFE
                     SPGPLHDIIVCWIDQHEVHKGGAKRLQLLVFELIRSGIFNPIAYVRQLIVSGMIDVIQ
                     PAVDPERRMRHHRILKQLPGCFVHETLEEAQLFGGDKLSEAVRTYSNERRLLLRELLV
                     EKGKYWNNLVLSDQKSKKISTSLSSVIFPRACNAKSNSKGPRKHTKSSVDIRELKERI
                     SALLQFPGMSCGVETPVRDEFQNSVKRSSGSVYSKMDQPEATPGCEDCRRAKRPKMND
                     EKSSCYQGNSPIASDEEDNWWIKKGSKTVESSLKVDPQIEITKQVPRGRQKMARKTQS
                     LAQLQAARIEGSQGASTSHVCDNKVSCPHHGPGVEGENQKVVDVFRTSTPVDMVSVGN
                     SLKQLQFVDKRSIAVWLTTAVRQLVEEPQKSSVRVGQFNRGAPVEEKNTIRWKLGADE
                     LYSILFLLDISLDLVSAGGRNLVTVPRNVENNMCEIGEAILVSSLRRYENILLSADLV
                     PEAMTALMNRAASLMSSNGKISGSAALVYTRYILKRYGSLPSVVEWHNNFKATSEKKL
                     LSELDHTRSGNGEYGNPLGVPAGVDNPDDYLRKKISIGGARPSRVGLSMRDVLQRHVE
                     EATHYLKKLIGTGTMKASLAEKNDDGYQVAQQIVVGLMDCIRQTGGAAQEGDPSLVSS
                     AVSAIINSVGLSVARITDFSLGNIYQNHPSGVDSSNIARYILRIHITCLCLLKEALGE
                     RQSRVFEIALATESSTALTGVFAPVKGSRGQHQLSPESYDSNANNSTIDMSNGTGKMA
                     LSRATKITAAVSALVIGSITHGVITLERIVGLLRLKDYLDFVQFVRRTKSSSNGSARS
                     MGASKVESPIEVYVHWFRLLVGNCKTVSEGLVLELVGESSVVAISRMQRMLPLKLVFP
                     PAYSIIAFVLWRPFVSNSNSNSSVHEDTHRLYQSLTMAFHDVIKHLPFRDVCFRDTQG
                     LYELIVADSTDAEFASVFESHGLDMHLKSVAFAPLRARLFLNSLIDCKVPSSGYSHEG
                     VSEAKNRHQGNGTKLVDKLVSVLDCLQPAKFHWQWVELRLLLNEQALAEKLENHDMPL
                     TDAIRSSCPTSEKPDASENEKNFIQILLTRLLVRPDAVPLFSEVVHLFGRSVEDSMLK
                     QAEWFLAGQDVLFGRKTIRQKLIIVGESKGLPTKPQFWKPWGWCNSSSSDHITANKAG
                     KKRKFEITSIEEGEVIEEGSGSRKVLLPRVLDENSPSVGYGITTERAFVQLVLPCIDQ
                     SSDESRSTFVNELVRQFSNIEQQLSSVTNRSTTSNKQMGTASSGSEISSNKGSTRKGL
                     RGGSPSLARRSSANTTDTSPPPSPAALRASMSLRLQFLLRLLPVICGEPSFKNTRHAL
                     ASTIVRLLGSRVVYEDYAVCSPRSELSKAETESTIDPSSMADLSSEVLFDRRELTGLS
                     TNFKKSLKVCLLGCKNELSRMQLPDTIRWRIQAAMPILLPSLRCSLSCQPHSVPPTAL
                     TLVQPSGSTAAAGTNQRNSPAISKSGTAAAQGKLKPTMLAPHQQQEADNTDVVDPWTL
                     LEDGTSSGLSSSNASNSSDMANLRATCWLKGAVRVRRTDLTYVGSVDDDS"
     exon            10352. .10384
                     /gene="AT4g00450"
                     /number=1
     intron          10385. .10431
                     /gene="AT4g00450"
                     /number=1
     exon            10432. .10545
                     /gene="AT4g00450"
                     /number=2
     intron          10546. .10604
                     /gene="AT4g00450"
                     /number=2
     exon            10605. .10709
                     /gene="AT4g00450"
                     /number=3
     intron          10710. .10848
                     /gene="AT4g00450"
```

```
                /number=3
exon            10849. .11018
                /gene="AT4g00450"
                /number=4
intron          11019. .11078
                /gene="AT4g00450"
                /number=4
exon            11079. .11423
                /gene="AT4g00450"
                /number=5
intron          11424. .11474
                /gene="AT4g00450"
                /number=5
exon            11475. .13674
                /gene="AT4g00450"
                /number=6
intron          13675. .13734
                /gene="AT4g00450"
                /number=6
exon            13735. .13823
                /gene="AT4g00450"
                /number=7
intron          13824. .13920
                /gene="AT4g00450"
                /number=7
exon            13921. .16530
                /gene="AT4g00450"
                /number=8
intron          16531. .16676
                /gene="AT4g00450"
                /number=8
exon            16677. .16862
                /gene="AT4g00450"
                /number=9
intron          16863. .16975
                /gene="AT4g00450"
                /number=9
exon            16976. .17042
                /gene="AT4g00450"
                /number=10
intron          17043. .17091
                /gene="AT4g00450"
                /number=10
exon            17092. .17547
                /gene="AT4g00450"
                /number=11
gene            18027. .19004
                /gene="AT4g00460"
gene            complement(join(18027. .18146,18234. .18578,18669. .19004))
                /gene="AT4g00460"
CDS             complement(join(18027. .18146,18234. .18578,18669. .19004))
                /gene="AT4g00460"
                /note="similarity to
                contains EST gb:AI993831.1"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB80855.1"
                /db_xref="GI:7267385"
                /db_xref="SPTREMBL:O23056"
                /translation="MVTSVFDPCLLQEALDSFQNTEFWYAEEGSLSMKSARSSTGSFR
                KVIVQRKEEKWWLPVPLVPSEGLSDKARKQLKNKRESTNQIHKAAMAINSSILSEMEI
                PDSYMTTLPKCGKSSVGDSIYRYMSGSGRFFPEQLLDCLNISSEHEAVQLADRVEASM
                YTWRRKSCLSNSKNSWNMVKDLMSTTERTDKNYVMAERAETLLFCLKQRYPELSQTSL
                DICKIQYNKDVGKAVLESYSRVLEGLAFNIVAWIDDVLYVDKTMRGSE"
exon            complement(18027. .18146)
                /gene="AT4g00460"
                /number=1
intron          complement(18147. .18233)
                /gene="AT4g00460"
                /number=1
exon            complement(18234. .18578)
                /gene="AT4g00460"
                /number=2
intron          complement(18579. .18668)
                /gene="AT4g00460"
                /number=2
exon            complement(18669. .19004)
                /gene="AT4g00460"
                /number=3
gene            19070. .19781
                /gene="AT4g00470"
gene            complement(join(19070. .19221,19270. .19386,19472. .19781))
                /gene="AT4g00470"
CDS             complement(join(<19070. .19221,19270. .19386,19472. .19781))
                /gene="AT4g00470"
                /note="similarity to
                contains EST gb:AA721856"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB80856.1"
                /db_xref="GI:7267386"
                /db_xref="SPTREMBL:Q9M163"
                /translation="MENLSNPDENDDHQSPRSIDQNDQSAVETPVYSTMSIDSFVYPR
                TCSETTSGFSDQIDETNSFCSEASPCNWPVLTESKSSKCLSGLEMQSNECLVVQEISE
                PELETMKERFAKLLLGEDMSGSGKGVCTAVTISNAITNLYDNCKNDTSFDNVCSYSVW
                TESEVRAVRNREESIVEERNELSFICMRLHCRI"

Alignment Scores:
Pred. No.:                3.94e-85      Length:          197975
Score:                    1116.50       Matches:         231
Percent Similarity:       68.38%        Conservative:    35
Best Local Similarity:    59.38%        Mismatches:      26
Query Match:              75.13%        Indels:          99
DB:                       8             Gaps:            3

US-10-723-061-18 (1-292) x ATCHRIV2 (1-197975)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              ||  :::|||||||||||||   |||||||||:::||||||:::||||||   :::|||
Db      67119 GCAATTGTAATAGATGGAAAGGCTGTGGCAAAAAAAGATTAGAGATGAAATCACAATTGAA 67060

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::||||||:::::||||||||||:::||||||||||||||||||||||||||
Db      67059 GTTTCAAGAATGAAGGAATCAATTGGTGTGATTCCTGGTTTAGCAGTAATCCTTGTTGGT 67000

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||   |||||||||||||||||||||||||||||||:::::||||||
Db      66999 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAGAAAGCTTGTGACTCCGTTGGA 66940

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::||||||   |||   ||||||||||||||:::||||||||||||
Db      66939 ATCAAATCGTTCGAAGTTCGTCTAGCTGAAGATTCATCAGAAGAAGAGGTGTTGAAATCT 66880

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              ::::::   ||||||   ||||||||||||||||||||||||||||||||||||||||
Db      66879 GTATCAGGATTCAATGATGATCCTTCTGTCCATGGAATCCTTGTTCAGTTGCCTCTGCCA 66820

Qy        101 His------------------------------------------------------------ 101

Db      66819 TC-GGTAAACAAAGTCTAATAACTTTCTTCGGAACTTAAATAAACAGTGAATTACAATAG 66761

Qy        102 --------------------HisMetAsnAspGluAsnIleLeuAsnAlaValSerIle 114
                                  ||||||:::::::::||||||||||||||||||||||||
Db      66760 TGTTTTGGGATGTTGTTGCAGCATATGGATGAACAGAACATACTGAATGCGGTTAGTATA 66701

Qy        115 GluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArg 134
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db      66700 GAGAAAGATGTTGACGGATTTCATCCGCTAAATATTGGACGGCTTGCCATGCGTGGAAGA 66641

Qy        135 AspProPhePheValProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGly 154
              :::|||   ||||||||||||||||||||||||||:::||||||||||||||||||
Db      66640 GAACCCTTATTCGTTCCTTGTACTCCAAAAGGATGCATTGAGTTGTTGCATAGATACAAC 66581

Qy        155 ValGluIleLysGlyLysArgAlaValValIleGlyArgSerAsnIleValGlyMetPro 174
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66580 ATTGAAATCAAAGGAAAGAGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGGTATGCCA 66521
```

```
Qy        175 AlaAlaLeuLeuLeuGln---------------------------------------- 180
              ||||||||||||||||||
Db      66520 GCTGCTCTTTTACTGCAGGTAATGATGATGATACCTTCTTTAAGATGCCGTTTTTTCTGG 66461

Qy        180 ------------------------------------------------------------ 180

Db      66460 TTGTGACATTCGAAAGTTAGTTACATGTTTGGTCTTTATTGTTGTGTTTTGTATTTGTAG 66401

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              :::   :::|||||||||||||||||:::||||||  ||||||  ||||||||||||||||
Db      66400 AGGGAGGATGCAACCGTTAGCATTATCCATTCAAGAACCAAGAACCCTGAAGAAATCACA 66341

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||:::|||:::|||||||||      |||:::|||||||||||||||
Db      66340 AGAGAAGCTGATATTATAATCTCAGCTGTTGGACAGCCAAACATGGTCAGAGGAAGCTGG 66281

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAsp------------ 236
              ||||||||||||||   :::|||||||||||||||||||||||||
Db      66280 ATAAAACCGGGCGCAGTCCTCATCGATGTTGGGATTAATCCTGTTGA-GGTGAGTGGAAT 66222

Qy        236 ------------------------------------------------------------ 236

Db      66221 TTTCTTTTTGACATATGGTCATGGCGTAATTCAGTGATATAATAAGTTTCAAATCAACCA 66162

Qy        237 ----------------------------------------AspProGluSerProArgGly 243
                                                       ||||||  :::  ||||||
Db      66161 GTATGCAACATTCATACATGTACATAATATGAATTGCAGGATCCAAGTGCTGCGCGTGGA 66102

Qy        244 TyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIleThr 263
              ||||||||||||||||||:::|||||||||||||||||||:::|||      ||||||
Db      66101 TATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAGTTGCATCAGCCATCACA 66042

Qy        264 ProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeuGlu 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66041 CCTGTTCCTGGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAACACTTTAACA 65982

Qy        284 SerAlaLysArgIleHisLysPheLys 292
              ||||||||||||||||||  |||:::
Db      65981 TCAGCTAAGAGGATTCACAACTTCCAG 65955
```

RESULT 9
AP006402/c

```
LOCUS       AP006402               92675 bp    DNA     linear   PLN 22-JUL-2003
DEFINITION  Lotus corniculatus var. japonicus genomic DNA, chromosome 1,
            clone:LjT34J23, TM0284, complete sequence.
ACCESSION   AP006402
VERSION     AP006402.1  GI:31581033
KEYWORDS    HTG.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Asamizu,E., Kato,T., Sato,S., Nakamura,Y., Kaneko,T. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. IV. Sequence
            Features and Mapping of seventy-three TAC clones which cover the
            7.5 Mb Regions of the Genome
  JOURNAL   DNA Res. (2003) In press
REFERENCE   2  (bases 1 to 92675)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2003) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .92675
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT34J23"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0284~synonym: Lotus japonicus"
ORIGIN
```

Alignment Scores:

Pred. No.:	3.89e-80	Length:	92675
Score:	1054.50	Matches:	228
Percent Similarity:	59.17%	Conservative:	30
Best Local Similarity:	52.29%	Mismatches:	34
Query Match:	70.96%	Indels:	145
DB:	8	Gaps:	3

US-10-723-061-18 (1-292) x AP006402 (1-92675)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              ||||||:::|||||||||||||||||||||||||||||||||||:::|||||||   |||
Db      19525 GCAAAGGTGATTGATGGAAAATTAGTGGCAAAGCAAATCAGAGATGAGATAACAGCTGAA 19466

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::::::::|||||||:::|||||||||||||||||||||||||||||||||||||||||
Db      19465 GTCTCTAGGATGAAAGAAGCCATTGGTGTGGTTCCAGGGTTAGCTGTAATCCTCGTTGGA 19406

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||  |||||||||||||||||||:::|||||||||||||||   |||
Db      19405 GATAGGAAGGACTCTGCAACTTATGTGCGTAACAAGCAGAAGGCTTGTGAAGCTGCTGGA 19346

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||   |||   |||       |||||||||||||:::|||:::|||||||||   :::
Db      19345 ATAAATTCTTTGGAAGCATTTTTGCCAGAAGATTCCACTGAAGAAGAAGTTTTGAATTAT 19286

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              |||:::   :::|||  ||||||||||||||||||||||||||||||||||||||||||
Db      19285 ATTTCAGGCTACAATGAAGATCCTTCAGTTCATGGCATTCTTGTTCAATTACCTTTACCT 19226

Qy        101 His---------------------------------------------------------- 101

Db      19225 TC-TGTATGCTTCTAGATCTCTAACTTTATTTTCTTTTCAGGGTAAAGTATTATACATTA 19167

Qy        101 ------------------------------------------------------------ 101

Db      19166 ATAACTTGTTCTCATAGCTAAGTTTACTGTTACGTTGACTATAACTATATAAGTAACTTA 19107

Qy        102 ----------------------------------------------HisMetAsnAsp 105
                                                           |||||||||:::
Db      19106 ATTTTTTTCATATGGAAGTTTAGTTAATTTTCCTTGCCATTGTGACAGCATATGAATGAG 19047

Qy        106 GluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHisProLeuAsn 125
              :::||||||||||||||||:::::|||||||||||||||||||||||||||||||||||
Db      19046 CAGAACATCTTGAATGCTATTACAATTGAGAAGGATGTAGATGGTTTTCATCCATTGAAT 18987

Qy        126 IleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThrProLysGly 145
              ||||||||||||||||:::   |||:::|||   ||||||||||||||||||||||||||
Db      18986 ATCGGTCGTCTTGCAATGCGTGATAGAGAACCCCTGTTCGTTCCCTGTACACCAAAGGGA 18927

Qy        146 CysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAlaValValIle 165
              |||:::|||||||||   ||||||||||  |||||||||  ||||||||||||||||||
Db      18926 TGCATAGAGTTACTCCGCAGATATGGTGTTTCTATCAAGGGAACGAGGGCTGTCGTGATT 18867

Qy        166 GlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln--------------- 180
              ||||||||||||||||||||||||||||||||||||||||||||
Db      18866 GGTCGGAGCAATATTGTAGGAATGCCAGCCGCTCTCTTGCTGCAAGTAATATCTCAATGG 18807

Qy        180 ------------------------------------------------------------ 180

Db      18806 CTTTAGTTAACAGTGAAATTTGGTTTTACTTCAAAGTGATTGTTTGCTCCTTTCTTTTTT 18747

Qy        181 ----------------------------------------LysAlaAsnAlaThrVa 186
                                                       :::   :::||||||||
Db      18746 GCATCTATGTTTCTACTTCTTTATATATCTCTTATGTGACTAGAGGGAAGATGCTACTGT 18687
```

```
Qy        186 lSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGlnAlaAspIleVa 206
              ||||:::|||||||||   |||      ||||||||||||   ||||||||||||||||||
Db      18686 CAGTGTCGTCCATTCAAAAACCCTAAACCCTGAAGAGATCATAAGACAAGCAGATATTGT 18627

Qy        206 lIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLysProGlyAlaAl 226
              ||||:::|||   |||      ::::::||||||||||||||||||||||||||||||||
Db      18626 CATCTCGGCTGCTGGGCAACCACACATGGTTAGGGGAAGCTGGATAAAACCTGGTGCTGT 18567

Qy        226 aIleIleAspValGlyIleAsnProValAsp---------------------------- 236
              ||||||||||||||||||||||||||||:::
Db      18566 CATTATCGACGTTGGAATCAACCCTGTAGAGGTGAATTACTTTTTTCGAGATCTCGTTGA 18507

Qy        236 ------------------------------------------------------------ 236

Db      18506 TAGTATGCTAGGGGTTTCTCTTGGTTATGTTTAGAGGAATAAATTCATAGTTAATACTTA 18447

Qy        237 ------------------------------------------------------------AspP 238
                                                                            ||||
Db      18446 AATTTCAATGTGAATGGTGTTTACTTAACCAATGCATATTTGAGTGTTTGTAACAGGATC 18387

Qy        238 roGluSerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysI 258
              ||   |||   ||||||||||||||:::|||||||||||||||||||||||||   |||:
Db      18386 CAAATAGCACTCGAGGTTATAGATTGATCGGAGATGTTTGCTATGAAGAAGCCATAAAAG 18327

Qy        258 leAlaGlyLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuL 278
              ::|||      :::||||||||||||||||||||||||||||||||||||||||||||||
Db      18326 TTGCCTCGGCTGTGACACCAGTGCCTGGAGGAGTAGGTCCAATGACCATAGCAATGCTTC 18267

Qy        278 euSerAsnThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              ||   |||||||||   ||||||||||||||||||||||:::
Db      18266 TCCAAAATACTCTTAACTCTGCAAAGAGGATCCACAAGTTTGAA 18223
```

RESULT 10
AP006092

LOCUS	AP006092 28160 bp DNA linear PLN 22-JUL-2003
DEFINITION	Lotus corniculatus var. japonicus genomic DNA, chromosome 4, clone:LjT21D16, TM0157b, complete sequence.
ACCESSION	AP006092
VERSION	AP006092.1 GI:29122731
KEYWORDS	HTG.
SOURCE	Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM	Lotus corniculatus var. japonicus Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
REFERENCE	1
AUTHORS	Kaneko,T., Asamizu,E., Kato,T., Sato,S., Nakamura,Y. and Tabata,S.
TITLE	Structural analysis of a Lotus japonicus genome. III. Sequence features and mapping of sixty-two TAC clones which cover the 6.7 Mb regions of the genome
JOURNAL	DNA Res. 10 (1), 27-33 (2003)
MEDLINE	22579290
PUBMED	12693552
REFERENCE	2 (bases 1 to 28160)
AUTHORS	Sato,S.
TITLE	Direct Submission
JOURNAL	Submitted (12-DEC-2002) Shusei Sato, Kazusa DNA Research Institute, Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp, URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337), Fax:81-438-52-3934)

```
FEATURES             Location/Qualifiers
     source          1. .28160
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="4"
                     /clone="LjT21D16"
                     /clone_lib="LjT library"
                     /note="TM0157b, a part of TAC clone:TM0157.~synonym: Lotus
                     japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                  4.6e-80     Length:          28160
Score:                      1046.50     Matches:           217
Percent Similarity:         84.30%      Conservative:       30
Best Local Similarity:      74.06%      Mismatches:         42
Query Match:                70.42%      Indels:              7
DB:                         8           Gaps:                3

US-10-723-061-18 (1-292) x AP006092 (1-28160)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              ||||||:::|||||||||||||||||||||   |||   |||   |||   |||   |||
Db      15379 GCTAAGGTGATTGATGGAAAATTAGTGGCAAATCAG---AGATGAGAACA-GCT---GAA 15431

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              ::::::::::||||||:::|||||||||||||||||||||||||||||||||||||||||
Db      15432 GTCTCTAGGATGAAAGAAGCCATTGGTGTGGTTCCAGGGTTAGCTGTAATCCTTGTTGGA 15491

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              |||||||||||||   |||||||||||||||||||:::|||||||||||||||   |||
Db      15492 GATAGGAAGGACTCCGCAACTTATGTGCGTAATAAGCAGAAGGCTTGTGAAGCTGCTGGA 15551

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||   |||   |||       ||||||||||||||:::|||::|||||||||   :::
Db      15552 ATAAATTCTTTGGAAGCATTTTTGCCGGAAGATTCCACTGAAGAAGAAGTTTTGAATTAT 15611

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              |||:::   :::|||   |||||||||||||||||||||||||   ||||||||||||
Db      15612 ATTTCAGGCTACAATGAAGATCCTTCAGTTCATGGCATTCTTGTTCCATTACCTTTACCT 15671

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||||||||:::::|||||||||||||||||:::::|||||||||||||||||||||||
Db      15672 TCTCATATGAATGAGCAGAACATCTTGAATGCTATTACAATTGAGAAGGATGTAGATGGT 15731

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||||||   |||||||||:::   |||:::|||       ||||||
Db      15732 TTTCATCCATTGAATATCGGTCATCTTGCAATGCGTGATAGAGAACCCCTGCTCGTTCCC 15791

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||||||||||||||||:::|||||||||   ||||||||||   ||||||||||||||
Db      15792 TGTACACCAAAGGGATGCATAGAGTTACTCCGCAGATATGGTGTTTCTATTAAGGGAAAG 15851

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db      15852 AGGGCTGTCGTGATT---GAGCAATA-TTTGTAGGAATGCCAGCTGCTCTCTTGCTGCAA 15907

Qy        181 LysAla-AsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleTh 200
              |||   :::|||||||||||||:::|||||||||   |||   |||||||||||||||
Db      15908 AAGGGAAGATGCTACTGTCAGTGTTGTCCATTCAAGAACCCTAAACCCTGAAGAGATCAT 15967

Qy        200 rArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTr 220
              :::|||||||||||||||:::|||||||||   ::::::|||||||||||||||||||
Db      15968 AAGTGAAGCAGATATTGTTATCTCGGCTGTTGGACAACCACACATGGTTAGGGGAAGCTG 16027

Qy        220 pIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSe 240
              |||||||||||||   ||||||||||||||||||||||||||||:::||||||   ||
Db      16028 GATAAAACCCGGTGTTGTCATTATCGACGTTGGAATCAACCCTGTAGAGGATCCGAATAG 16087

Qy        240 rProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGl 260
              ||||:::|||||||||||||||||||||:::|||||||||   |||:::|||
Db      16088 CCCTCAAGGTTATAGATTGGTCGGAGATGTTTGCTTTGAAGAAGCAATAAAAGTTGCCTC 16147

Qy        260 yLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAs 280
              :::|||   ||||||||||||||||||||||||||||||||||||||||||   ||
Db      16148 AGCTGTGACACA-GTGCCTGGAGGAGTAGGTCCAATGACCATAGCAATGCTTCTCCAAAA 16206

Qy        280 nThrLeuGluSerAlaLysArgIleHisLysPheLys 292
              |||||||   ||||||||||||||||||||||:::
Db      16207 TACTCTTAACTCTGCAAAGAGGATCCACAAGTTTGAA 16243
```

```
RESULT 11
AY085538
LOCUS       AY085538               1182 bp    mRNA    linear   PLN 14-APR-2003
DEFINITION  Arabidopsis thaliana clone 156450 mRNA, complete sequence.
ACCESSION   AY085538
VERSION     AY085538.1  GI:21404248
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1182)
  AUTHORS   Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N.,
            Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
  TITLE     Full-length messenger RNA sequences greatly improve genome
            annotation
  JOURNAL   Genome Biol. 3 (6), RESEARCH0029 (2002)
  MEDLINE   22088475
   PUBMED   12093376
REFERENCE   2  (bases 1 to 1182)
  AUTHORS   Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
  TITLE     Full-Length cDNA from Arabidopsis thaliana
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1182)
  AUTHORS   Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
            Malibu, CA 90265, USA
COMMENT     This clone sequence is one of 5,000 Ceres full-length cDNAs made
            available to TIGR and Genbank. The following quality assessment of
            this set was done by comparison with known proteins: two percent of
            the clones are estimated to be 5'-truncated; less than one percent
            are 3'-truncated; approximately two percent represent alternative
            splice variants, including unspliced introns and spliced exons; one
            percent may contain premature stop codons; five percent may have
            frame shifts in a coding region. A sequence is considered to be
            5'-truncated if it lacks the translation initiation start (ATG). A
            sequence is considered to be 3'-truncated if it lacks the
            C-terminal end of the encoded protein. Please note that these cDNA
            sequences are derived from the Ws or LAer ecotypes and therefore
            may contain polymorphisms when compared to sequences from Col-0.
            Genset carried out the library production and sequencing of the
            full-length clones. Ceres, Inc. carried out the clustering of the
            5' sequences, selection of clones, and sequence assembly.
FEATURES             Location/Qualifiers
     source          1. .1182
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="156450"
     CDS             90. .989
                     /codon_start=1
                     /product="5,10-methylenetetrahydrofolate
                     dehydrogenase:5,10-methenyltetrahydrofolate
                     cyclohydrolase, putative"
                     /protein_id="AAM62762.1"
                     /db_xref="GI:21553669"
                     /translation="MASSSDHTAKIIDGKAIAHTIRSEIAEEVRGLSEKHGKVPGLAV
                     VIVGSRKDSQTYVNTKRKACAEVGIKSFDVGLPEEVSEADLISKVHELNSNPDVHGIL
                     VQLPLPKHINEENILGAISIDKDVDGFHPLNIGKLAMKGREPLFLPCTPKGCLELLAR
                     SGVKIKGQRAVVVGRSNIVGLPVSLLLLKADATVTTVHSHTKDPEAIIREADIVIAAC
                     GQAHMIKGNWIKPGAAVIDVGTNAVSDPSKKSGYRLVGDVDFAEASKVAGFITPVPGG
                     VGPMTVAMLLRNTVDGAKRVFGE"
ORIGIN

Alignment Scores:
Pred. No.:                1.01e-80        Length:          1182
Score:                    1035.00         Matches:          194
Percent Similarity:       82.99%          Conservative:     45
Best Local Similarity:    67.36%          Mismatches:       49

Query Match:              69.65%          Indels:           0
DB:                       8               Gaps:             0

US-10-723-061-18 (1-292) x AY085538 (1-1182)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||||||||||||||   :::|||      ||||||  |||||||||    |||
Db        114 GCGAAGATAATCGACGGCAAGGCGATTGCTCATACCATCAGATCAGAGATCGCCGAGGAA 173

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::      :::   :::      |||   ||||||||||||||||||:::::|||||
Db        174 GTTCGCGGTCTATCTGAGAAACACGGCAAGGTCCCAGGACTAGCTGTAGTTATTGTTGGG 233

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||||||      |||:::|||||||||      ||||||
Db        234 AGCCGAAAGGATTCACAGACCTATGTGAATACGAAGAGGAAAGCGTGCGCTGAGGTTGGG 293

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::::|||   |||||||||:::   ||||||   :::::::::
Db        294 ATTAAGTCATTTGACGTGGGCCTACCAGAGGAAGTTTCTGAAGCTGATCTTATTAGCAAA 353

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::      ||||||:::|||  ||||||||||||||||||||||||||||||||||||
Db        354 GTTCATGAACTAAATTCAAATCCGGATGTCCATGGCATATTAGTTCAACTCCCATTGCCG 413

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||:::|||:::|||||||||||||   |||:::|||||||:::|||||||||||||||
Db        414 AAACATATTAATGAGGAGAATATATTGGGTGCAATCAGCATTGATAAAGATGTCGATGGC 473

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||:::|||||||||:::|||||:::|||   |||:::|||
Db        474 TTCCATCCTTTGAATATTGGTAAGCTAGCCATGAAAGGCAGAGAACCCCTCTTCCTTCCT 533

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||:::|||||||||   |||   ||||||:::||||||||||:::
Db        534 TGCACCCCAAAGGGATGTTTGGAACTCCTAGCTAGAAGCGGCGTAAAGATAAAGGGGCAA 593

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||:::|||||||||||||||||||:::|||   :::||||||||||
Db        594 CGAGCAGTTGTTGTAGGTCGGAGTAACATTGTTGGTTTGCCCGTTTCACTTCTTTTGCTC 653

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::||||||||||:::   |||||||||:::||||||   ||||||   |||
Db        654 AAGGCTGATGCTACTGTCACAACTGTACATTCTCACACCAAGGATCCTGAGGCTATCATA 713

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||   |||   |||:::::::::::::|||:::|||
Db        714 CGGGAAGCTGACATTGTTATTGCTGCATGCGGACAAGCCCACATGATTAAGGGCAACTGG 773

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||:::|||||||||||   |||   |||   ||||||
Db        774 ATAAAGCCAGGGGCTGCAGTAATTGATGTTGGAACTAATGCAGTCAGCGACCCGAGCAAG 833

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||||||||||   :::   ||||||||||||:::||||||
Db        834 AAATCAGGATACCGGTTGGTTGGAGATGTTGATTTCGCAGAAGCTTCAAAAGTTGCAGGT 893

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||:::||||||||||||   |||
Db        894 TTCATAACTCCGGTCCCTGGTGGTGTAGGCCCAATGACAGTGGCAATGCTTCTCAGGAAC 953

Qy        281 ThrLeuGluSerAlaLysArgIle 288
              |||::::::   |||||||||:::
Db        954 ACCGTAGACGGTGCCAAGCGTGTC 977

RESULT 12
AF030516
LOCUS       AF030516               1219 bp    mRNA    linear   PLN 05-JAN-1999
DEFINITION  Pisum sativum 5,10-methylenetetrahydrofolate
            dehydrogenase-5,10-methenyltetrahydrofolate cyclohydrolase mRNA,
            complete cds.
ACCESSION   AF030516
VERSION     AF030516.1  GI:4103986
```

```
KEYWORDS    .
SOURCE      Pisum sativum (pea)
  ORGANISM  Pisum sativum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Pisum.
REFERENCE   1  (bases 1 to 1219)
  AUTHORS   Chen,L., Nargang,N.E. and Cossins,E.A.
  TITLE     a cDNA sequence encoding pea leaf cytosolic bifunctional
            5,10-methylenetetrahydrofolate
            dehydrogenase-5,10-methenyltetrahydrofolate cyclohydrolase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1219)
  AUTHORS   Chen,L., Nargang,N.E. and Cossins,E.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Chemistry and Biochemistry, University of
            Texas at Austin, Austin, TX 78705, USA
FEATURES             Location/Qualifiers
     source          1. .1219
                     /organism="Pisum sativum"
                     /mol_type="mRNA"
                     /db_xref="taxon:3888"
     CDS             68. .952
                     /note="bifunctional protein; 31.3 kDa protein"
                     /codon_start=1
                     /product="5,10-methylenetetrahydrofolate
                     dehydrogenase-5,10-methenyltetrahydrofolate
                     cyclohydrolase"
                     /protein_id="AAD01907.1"
                     /db_xref="GI:4103987"
                     /translation="MATVIDGKAVAQTIRSEIADEVRLLSQKYGKVPGLAVVIVGNRK
                     DSQSYVGMKRKACAELGIKSFDIDLPEDASEAEIIKNVHELNANPDVHGILVQLPLPK
                     HVNEEKVLTEISISKDVDGFHPLNIGKLAMKGRDPLFLPCTPKACLELLSRSGVSIKG
                     KKAVVVGRSNIVGLPASLLLLKADATVTIVHSHTSQPETIIREADIVIAAAGQAKMIK
                     GSWIKPGAAVIDVGTNSVDDPTRKSGYRLVGDVDFEEASKVAGWITPVPGGVGPMTVT
                     MLLKNTLEGAKRTIEQNN"
ORIGIN

Alignment Scores:
Pred. No.:                1.28e-80        Length:          1219
Score:                    1034.00         Matches:         193
Percent Similarity:       82.93%          Conservative:    45
Best Local Similarity:    67.25%          Mismatches:      49
Query Match:              69.58%          Indels:          0
DB:                       8               Gaps:            0

US-10-723-061-18 (1-292) x AF030516 (1-1219)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::|||||||||||||   ||||||:::   ||||||   |||||||||   |||
Db         71 GCCACCGTAATCGACGGCAAAGCCGTTGCACAAACTATCCGATCTGAAATCGCCGACGAG 130

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::      :::            |||   ||||||||||||||||||:::::||||||
Db        131 GTTCGTCTTCTCTCACAAAAGTACGGCAAGGTTCCTGGACTTGCAGTGGTGATAGTAGGA 190

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              :::|||||||||||||:::||||||      |||:::|||||||||      :::|||
Db        191 AACCGAAAGGACTCTCAAAGCTACGTTGGAATGAAGAGAAAAGCGTGTGCTGAATTGGGA 250

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::::::::::::|||||||||||:::||||||   |||:::::|||:::
Db        251 ATCAAATCCTTCGATATCGACCTTCCGGAGGATGCTTCTGAAGCTGAAATCATAAAAAAT 310

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::      |||::::::|||   ||||||||||||||||||||||||||||||||||||
Db        311 GTTCACGAATTGAACGCTAACCCCGATGTACATGGTATATTGGTTCAACTTCCATTGCCT 370

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                 |||:::|||:::|||   :::|||      :::|||||||   ||||||||||||||
Db        371 AAGCACGTAAATGAAGAGAAAGTTTTGACTGAAATCAGCATTTCAAAGGATGTAGATGGC 430

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||||||:::|||||||||:::|||||||||||   |||:::|||
Db        431 TTCCATCCCTTGAACATTGGCAAGCTTGCAATGAAAGGAAGAGACCCTCTGTTTCTTCCA 490

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||   |||:::|||||||||   |||   ||||||   ||||||||||||
Db        491 TGTACTCCCAAGGCATGTCTTGAACTATTATCACGAAGTGGTGTAAGTATAAAGGGGAAA 550

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              :::|||||||||:::|||||||||||||||||||||:::||||||:::|||||||||
Db        551 AAGGCTGTTGTGGTTGGTAGAAGCAACATAGTTGGATTACCAGCTTCATTGCTGCTTCTG 610

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::||||||||||:::|||||||||||||:::|||   :::||||||   |||
Db        611 AAAGCAGATGCTACAGTTACCATTGTTCATTCACACACAAGTCAACCAGAAACTATCATT 670

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||||   |||   |||   :::::::::|||||||||
Db        671 CGTGAAGCAGATATTGTTATTGCAGCAGGACAGGCAAAGATGATCAAGGGAAGCTGG 730

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              |||||||||||||||||||:::||||||||||||   |||   ||||||||||||
Db        731 ATAAAACCAGGAGCTGCAGTGATAGATGTTGGCACAAATTCTGTGGATGACCCAACTAGG 790

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
                  |||||||||||||||||||||||||   :::|||||||||||||||:::||||||
Db        791 AAATCAGGTTATAGACTTGTTGGAGATGTAGATTTTGAGGAAGCATCTAAAGTTGCTGGT 850

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||:::   |||||||||   |||
Db        851 TGGATTACTCCTGTTCCTGGTGGTGTGGGTCCAATGACAGTCACAATGTTGCTGAAGAAT 910

Qy        281 ThrLeuGluSerAlaLysArg 287
              |||||||||   |||||||||
Db        911 ACTTTGGAGGGCGCTAAACGC 931

RESULT 13
PSA011589
LOCUS       PSA011589               1219 bp    mRNA    linear   PLN 29-SEP-1999
DEFINITION  Pisum sativum mRNA for bifunctional 5,10-methylene-
            tetrahydrofolate dehydrogenase and 5,10-methenyltetrahydrofolate
            cyclohydrolase protein.
ACCESSION   AJ011589
VERSION     AJ011589.1  GI:6002382
KEYWORDS    5,10-methenyltetrahydrofolate cyclohydrolase;
            5,10-methylene-tetrahydrofolate dehydrogenase; bifunctional enzyme.
SOURCE      Pisum sativum (pea)
  ORGANISM  Pisum sativum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Pisum.
REFERENCE   1
  AUTHORS   Chen,L., Nargang,F.E. and Cossins,E.A.
  TITLE     Isolation and sequencing of a cDNA encoding pea cytosolic
            bifunctional 5,10-methylenetetrahydrofolate
            dehydrogenase:5,10-methenyltetrahydrofolate cyclohydrolase.
            Comparisons with the yeast, mammalian, and bacterial proteins
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1219)
  AUTHORS   Chen,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-1998) Chen L., Biochemistry and Chemistry,
            University of Texas at Austin, Speedway and 24th St., Austin, Texas
            78731, USA
FEATURES             Location/Qualifiers
     source          1. .1219
                     /organism="Pisum sativum"
                     /mol_type="mRNA"
                     /db_xref="taxon:3888"
     CDS             68. .952
                     /EC_number="1.5.1.5"
                     /note="bifunctional enzyme"
```

```
                     /codon_start=1
                     /evidence=experimental
                     /product="5,10-methylenetetrahydrofolate dehydrogenase:
                     5,10-methenyltetrahydrofolate cyclohydrolase"
                     /protein_id="CAB56756.1"
                     /db_xref="GI:6002383"
                     /db_xref="GOA:Q9ZTV0"
                     /db_xref="SPTREMBL:Q9ZTV0"
                     /translation="MATVIDGKAVAQTIRSEIADEVRLLSQKYGKVPGLAVVIVGNRK
                     DSQSYVGMKRKACAELGIKSFDIDLPEDASEAEIIKNVHELNANPDVHGILVQLPLPK
                     HVNEEKVLTEISISKDVDGFHPLNIGKLAMKGRDPLFLPCTPKACLELLSRSGVSIKG
                     KKAVVVGRSNIVGLPASLLLLKADATVTIVHSHTSQPETIIREADIVIAAAGQAKMIK
                     GSWIKPGAAVIDVGTNSVDDPTRKSGYRLVGDVDFEEASKVAGWITPVPGGVGPMTVT
                     MLLKNTLEGAKRTIEQNN"
ORIGIN

Alignment Scores:
Pred. No.:              1.28e-80        Length:         1219
Score:                  1034.00         Matches:        193
Percent Similarity:     82.93%          Conservative:   45
Best Local Similarity:  67.25%          Mismatches:     49
Query Match:            69.58%          Indels:         0
DB:                     8               Gaps:           0

US-10-723-061-18 (1-292) x PSA011589 (1-1219)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||   :::|||||||||||||||   ||||||:::   ||||||   |||||||||   |||
Db         71 GCCACCGTAATCGACGGCAAAGCCGTTGCACAAACTATCCGATCTGAAATCGCCGACGAG 130

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::      :::              |||   ||||||||||||||||||:::::||||||
Db        131 GTTCGTCTTCTCTCACAAAAGTACGGCAAGGTTCCTGGACTTGCAGTGGTGATAGTAGGA 190

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              :::|||||||||||||||:::||||||     |||:::||||||||      :::|||
Db        191 AACCGAAAGGACTCTCAAAGCTACGTTGGAATGAAGAGAAAAGCGTGTGCTGAATTGGGA 250

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||:::::::::::::||||||||||||:::||||||   |||::::::|||:::
Db        251 ATCAAATCCTTCGATATCGACCTTCCGGAGGATGCTTCTGAAGCTGAAATCATAAAAAAT 310

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::         |||::::::|||   ||||||||||||||||||||||||||||||||||||
Db        311 GTTCACGAATTGAACGCTAACCCCGATGTACATGGTATATTGGTTCAACTTCCATTGCCT 370

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
                 |||:::|||:::|||   :::|||      :::||||||   ||||||||||||||||
Db        371 AAGCACGTAAATGAAGAGAAAGTTTTGACTGAAATCAGCATTTCAAAGGATGTAGATGGC 430

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||:::||||||||:::|||||||||   |||:::|||
Db        431 TTCCATCCCTTGAACATTGGCAAGCTTGCAATGAAAGGAAGAGACCCTCTGTTTCTTCCA 490

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||   |||:::|||||||||   |||   ||||||   ||||||||||||
Db        491 TGTACTCCCAAGGCATGTCTTGAACTATTATCACGAAGTGGTGTAAGTATAAAGGGGAAA 550

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              :::||||||||||:::|||||||||||||||||||||:::||||||:::||||||||||
Db        551 AAGGCTGTTGTGGTTGGTAGAAGCAACATAGTTGGATTACCAGCTTCATTGCTGCTTCTG 610

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::||||||||:::|||||||||||:::|||   :::||||||   |||
Db        611 AAAGCAGATGCTACAGTTACCATTGTTCATTCACACACAAGTCAACCAGAAACTATCATT 670

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||   |||   |||   ::::::::|||||||||
Db        671 CGTGAAGCAGATATTGTTATTGCAGCAGGACAGGCAAAGATGATCAAGGGAAGCTGG 730

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              |||||||||||||||||:::||||||||||||   |||   ||||||||||||
Db        731 ATAAAACCAGGAGCTGCAGTGATAGATGTTGGCACAAATTCTGTGGATGACCCAACTAGG 790

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||||||||||   :::|||||||||||||||||:::||||||
Db        791 AAATCAGGTTATAGACTTGTTGGAGATGTAGATTTTGAGGAAGCATCTAAAGTTGCTGGT 850

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||:::   |||||||||   |||
Db        851 TGGATTACTCCTGTTCCTGGTGGTGTGGGTCCAATGACAGTCACAATGTTGCTGAAGAAT 910

Qy        281 ThrLeuGluSerAlaLysArg 287
              |||||||||   |||||||||
Db        911 ACTTTGGAGGGCGCTAAACGC 931

RESULT 14
BT004612
LOCUS       BT004612                900 bp    mRNA    linear   PLN 19-FEB-2003
DEFINITION  Arabidopsis thaliana At3g12290 gene, complete cds.
ACCESSION   BT004612
VERSION     BT004612.1  GI:28416654
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 900)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 900)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2003) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAFL cDNAs: Cheuk,R.,
            Chen,H., Kim,C.J., Shinn,P., Bowser,L., Chan, M.M., Chang,C.M.,
            Dale,J.M., Hsuan,V.W., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera, C.S., Palm,C.J.,
            Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
            Yamada,K., Yu,G., Yuan,S., Davis,R.W.,  Theologis,A., and
            Ecker,J.R.

            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES                Location/Qualifiers
     source             1. .900
                        /organism="Arabidopsis thaliana"
                        /mol_type="mRNA"
                        /db_xref="taxon:3702"
```

```
                     /chromosome="3"
                     /clone="U22770"
                     /ecotype="Columbia"
                     /note="This clone is in pUNI 51"
     CDS             1. .900
                     /note="'5,10-methylenetetrahydrofolate
                     dehydrogenase:5,10-methenyltetrahydrofolate
                     cyclohydrolase, putative'"
                     /codon_start=1
                     /product="At3g12290"
                     /protein_id="AAO42858.1"
                     /db_xref="GI:28416655"
                     /translation="MASSSDHTAKIIDGKAIAHTIRSEIAEEVRGLSEKHGKVPGLAV
                     VIVGSRKDSQTYVNTKRKACAEVGIKSFDVGLPEEVSEADLISKVHELNSNPDVHGIL
                     VQLPLPKHINEEHILGAISIDKDVDGFHPLNIGKLAMKGREPLFLPCTPKGCLELLAR
                     SGVKIKGQRAVVVGRSNIVGLPVSLLLLKADATVTTVHSHTKDPEAIIREADIVIAAC
                     GQAHMIKGNWIKPGAAVIDVGTNAVSDPSKKSGYRLVGDVDFAEASKVAGFITPVPGG
                     VGPMTVAMLLRNTVDGAKRVFGE"
ORIGIN

Alignment Scores:
Pred. No.:                1.98e-80      Length:          900
Score:                    1030.00       Matches:         193
Percent Similarity:       82.99%        Conservative:    46
Best Local Similarity:    67.01%        Mismatches:      49
Query Match:              69.31%        Indels:          0
DB:                       8             Gaps:            0

US-10-723-061-18 (1-292) x BT004612 (1-900)

Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||||||||||||||||||   :::|||     ||||||   |||||||||   |||
Db         25 GCGAAGATAATCGACGGCAAGGCGATTGCTCATACCATCAGATCAGAGATCGCCGAGGAA 84

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::     :::   :::     |||   ||||||||||||||||||:::::||||||
Db         85 GTTCGCGGTCTATCTGAGAAACACGGCAAGGTCCCAGGACTAGCTGTAGTTATTGTTGGG 144

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||     |||:::|||||||||     ||||||
Db        145 AGCCGAAAGGATTCACAGACCTATGTGAATACGAAGAGGAAAGCGTGCGCTGAGGTTGGG 204

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||||||||::::::|||   |||||||||:::   ||||||   :::::::::
Db        205 ATTAAGTCATTTGACGTGGGCCTACCAGAGGAAGTTTCTGAAGCTGATCTTATTAGCAAA 264

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::     ||||||:::|||   ||||||||||||||||||||||||||||||||
Db        265 GTTCATGAACTAAATTCAAATCCGGATGTCCATGGCATATTAGTTCAACTCCCATTGCCG 324

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||:::|||:::|||:::||||||   |||:::|||||||:::|||||||||||||||
Db        325 AAACATATTAATGAGGAGCATATATTGGGTGCAATCAGCATTGATAAAGATGTCGATGGC 384

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              ||||||||||||||||||:::|||||||||:::||||||:::|||   |||:::|||
Db        385 TTCCATCCTTTGAATATTGGTAAGCTAGCCATGAAAGGCAGAGAACCCCTCTTCCTTCCT 444

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              ||||||||||||||||||:::|||||||||   |||   ||||||:::||||||||:::
Db        445 TGCACCCCAAAGGGATGTTTGGAACTCCTAGCTAGAAGCGGCGTAAAGATAAAGGGGCAA 504

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              ||||||||||||:::|||||||||||||||||||:::|||   :::|||||||||
Db        505 CGAGCAGTTGTTGTAGGTCGGAGTAACATTGTTGGTTTGCCCGTTTCACTTCTTTTGCTC 564

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::|||||||||:::   |||||||||:::||||||   ||||||   |||
Db        565 AAGGCTGATGCTACTGTCACAACTGTACATTCTCACACCAAGGATCCTGAGGCTATCATA 624

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||   |||   |||:::::::::::|||:::|||
Db        625 CGGGAAGCTGACATTGTTATTGCTGCATGCGGACAAGCCCACATGATTAAGGGCAACTGG 684

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              ||||||||||||||||||:::|||||||||||   |||   |||   ||||||
Db        685 ATAAAGCCAGGGGCTGCAGTAATTGATGTTGGAACTAATGCAGTCAGCGACCCGAGCAAG 744

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||||   :::   ||||||||||||:::||||||
Db        745 AAATCAGGATACCGGTTGGTTGGAGATGTTGATTTCGCAGAAGCTTCAAAAGTTGCAGGT 804

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              ||||||||||||||||||||||||||||||||||||:::|||||||||||   |||
Db        805 TTCATAACTCCGGTCCCTGGTGGTGTAGGCCCAATGACAGTGGCAATGCTTCTCAGGAAC 864

Qy        281 ThrLeuGluSerAlaLysArgIle 288
              |||::::::   ||||||||:::
Db        865 ACCGTAGACGGTGCCAAGCGTGTC 888

RESULT 15
AK103505
LOCUS       AK103505               1118 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:J033131D02, full
            insert sequence.
ACCESSION   AK103505
VERSION     AK103505.1  GI:32988714
KEYWORDS    FLI_CDNA; CAP trapper.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
  TITLE     Collection, mapping, and annotation of over 28,000 cDNA clones from
            japonica rice
  JOURNAL   Science 301 (5631), 376-379 (2003)
  MEDLINE   22752273
   PUBMED   12869764
REFERENCE   2  (bases 1 to 1118)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
            Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
            Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
            Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
            Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
            Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
            Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
            Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
            Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
            Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K.,
            Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
            Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
            Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
            Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
            Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
            Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
            Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
```

Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.

TITLE Direct Submission

JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers

source 1. .1118
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J033131D02"

ORIGIN

Alignment Scores:

Pred. No.:	9.45e-79	Length:	1118
Score:	1012.00	Matches:	191
Percent Similarity:	80.84%	Conservative:	41
Best Local Similarity:	66.55%	Mismatches:	55
Query Match:	68.10%	Indels:	0
DB:	8	Gaps:	0

US-10-723-061-18 (1-292) x AK103505 (1-1118)

```
Qy          1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
              |||:::|||||||||||||||   ||||||      ||||||   |||:::|||   :::
Db        117 GCGCAGATCATCGACGGGAAGGCGGTCGCCGCCGACATCCGCCGCGAGGTCGCCGCCGAC 176

Qy         21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
              :::|||   :::      |||      :::||||||||||||||||||:::::||||||
Db        177 GTCGCCGCCCTCTCCTCCGCCCACAACCTCGTGCCCGGGCTGGCGGTGGTGATCGTGGGG 236

Qy         41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
              ||||||||||||||||||||||||||:::   |||:::|||||||||      ||||||
Db        237 AGCAGGAAGGACTCGCAGACGTACGTGCAGATGAAGCGCAAGGCGTGCGCCGAGGTCGGC 296

Qy         61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
              |||:::|||   :::|||:::|||   ||||||   ||||||      ::::::
Db        297 ATCCGCTCCGTCGACGTCGACCTCGCCGAGGACATCTCCGAGGCGGCGCTCGTCGCCGAG 356

Qy         81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
              :::         |||:::|||||||:::||||||||||||||||||||||||||||||||
```

```
Db        357 GTCCACCGCCTCAACGCCGACCCAGCCGTCCATGGCATACTTGTCCAGCTTCCATTGCCA 416

Qy        101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
              |||:::|||:::|||   |||||||||   :::|||:::|||||||||||||||||||
Db        417 AAGCATATTAATGAGGAAAAGATATTGAACGAGATCAGCTTAGAGAAAGATGTTGATGGG 476

Qy        121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
              |||||||||||||||||||:::|||||||:::||||||||||   |||:::|||
Db        477 TTCCATCCTCTGAATATTGGCAAGCTTGCAATGAAAGGCAGAGACCCACTGTTCCTACCA 536

Qy        141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
              |||||||||||||||||||||||||||||   |||   ||||||   |||   ||||||
Db        537 TGCACGCCGAAGGGATGCATGGAGCTCCTAACACGGAGCGGAGTTACCATCAATGGGAAG 596

Qy        161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLeuGln 180
              |||||||||||:::|||||||||||||||||||:::|||   :::|||||||||
Db        597 CGAGCGGTCGTGGTTGGCCGCAGCAACATTGTCGGGCTACCTGTATCCCTGCTTCTTCTG 656

Qy        181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
              ||||||:::|||||||||||||||||||||||   |||      ||||||   |||
Db        657 AAGGCGGATGCGACCGTATCCATTGTTCACTCCCGGACCCCAAACCCTGAAAGTATTGTC 716

Qy        201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
              |||:::|||||||||||||||||||||   |||   |||   :::::::::|||   |||
Db        717 CGTGAAGCAGACATTGTCATTGCAGCGGCTGGCCAGGCTATGATGATCAAAGGAGACTGG 776

Qy        221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
              |||||||||||||||||:::|||||||||||   |||   :::   ||||||
Db        777 ATCAAACCAGGCGCTGCTGTCATCGACGTTGGGACGAACTCCATCAGTGACCCAACCAGG 836

Qy        241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
              ||||||||||||||||||||||||   :::   |||   ||||||:::||||||
Db        837 AAATCGGGGTACAGACTCGTCGGCGATGTGGATTTCGCAGAGGTGAGCAAGGTTGCTGGT 896

Qy        261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
              :::|||||||||||||||||||||||||||||||||:::||||||||||||   |||
Db        897 CACCTGACTCCAGTCCCAGGTGGCGTTGGCCCCATGACGGTGGCGATGTTGCTGAAGAAC 956

Qy        281 ThrLeuGluSerAlaLysArg 287
              |||::::::   |||||||||
Db        957 ACGGTGGATGGAGCGAAACGT 977
```

Search completed: July 24, 2004, 08:50:28
Job time : 5525 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 24, 2004, 06:09:35 ; Search time 84 Seconds
(without alignments)
1096.801 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1238	83.3	360	10	O65271	O65271 arabidopsis
2	1035	69.7	299	10	Q8LEA3	Q8lea3 arabidopsis
3	1034	69.6	294	10	Q9ZTV0	Q9ztv0 pisum sativ
4	1030	69.3	299	10	Q9LHH7	Q9lhh7 arabidopsis
5	1017	68.4	350	10	Q9ZVI8	Q9zvi8 arabidopsis
6	1009	67.9	310	10	O65269	O65269 arabidopsis
7	992.5	66.8	291	10	Q9FRB2	Q9frb2 oryza sativ
8	792	53.3	160	10	Q8W527	Q8w527 zea mays (m
9	765.5	51.5	937	3	Q9HFV0	Q9hfv0 yarrowia li
10	763.5	51.4	284	16	Q92BZ4	Q92bz4 listeria in
11	755.5	50.8	284	16	Q8Y7C5	Q8y7c5 listeria mo
12	749.5	50.4	937	3	Q8WZJ7	Q8wzj7 schizosacch
13	743.5	50.0	284	16	Q99YX2	Q99yx2 streptococc
14	743.5	50.0	284	16	Q8K6S8	Q8k6s8 streptococc
15	741	49.9	299	16	Q8FVP6	Q8fvp6 brucella su
16	741	49.9	319	16	Q8YCL8	Q8ycl8 brucella me
17	740.5	49.8	934	13	Q7ZXF4	Q7zxf4 xenopus lae
18	738	49.7	725	11	Q8BXX7	Q8bxx7 mus musculu
19	738	49.7	935	11	Q8R013	Q8r013 mus musculu
20	730	49.1	344	16	Q9A8Y0	Q9a8y0 caulobacter
21	724	48.7	285	16	Q97RI9	Q97ri9 streptococc
22	721	48.5	284	16	Q8E6M3	Q8e6m3 streptococc
23	721	48.5	284	16	Q8E168	Q8e168 streptococc
24	716	48.2	279	16	Q9K966	Q9k966 bacillus ha
25	715	48.1	299	16	Q989A7	Q989a7 rhizobium l
26	715	48.1	306	16	Q8DQD3	Q8dqd3 streptococc
27	714	48.0	284	16	Q8DVC1	Q8dvc1 streptococc
28	709.5	47.7	285	16	Q8EQ43	Q8eq43 oceanobacil
29	707.5	47.6	294	16	Q9CH85	Q9ch85 lactococcus
30	704.5	47.4	281	16	Q836W7	Q836w7 enterococcu
31	704	47.4	284	16	Q9JWI9	Q9jwi9 neisseria m
32	703	47.3	932	5	Q27772	Q27772 s c-1-tetra
33	701.5	47.2	291	16	Q7VVW3	Q7vvw3 bordetella
34	699.5	47.1	283	16	Q7WJG1	Q7wjg1 bordetella
35	699.5	47.1	283	16	Q7WAC0	Q7wac0 bordetella
36	699.5	47.1	283	16	Q7VZC8	Q7vzc8 bordetella
37	698.5	47.0	291	16	Q7W4Z7	Q7w4z7 bordetella
38	694	46.7	294	16	Q89WX9	Q89wx9 bradyrhizob
39	689	46.4	282	16	Q9PP68	Q9pp68 campylobact
40	689	46.4	299	16	Q92RY3	Q92ry3 rhizobium m
41	689	46.4	327	16	Q7UNN9	Q7unn9 rhodopirell
42	685.5	46.1	289	16	Q8XZ10	Q8xz10 ralstonia s
43	685	46.1	284	16	Q8EG21	Q8eg21 shewanella
44	684.5	46.1	286	16	Q818R5	Q818r5 bacillus ce
45	680.5	45.8	289	16	Q88WM8	Q88wm8 lactobacill

ALIGNMENTS

RESULT 1
O65271
ID O65271 PRELIMINARY; PRT; 360 AA.
AC O65271;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE F6N23.26 protein (Putative tetrahydrofolate synthase) (Hypothetical
DE protein).
GN F6N23.26 OR AT4G00620.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA WASHU;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Geisel C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Waterston R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.

```
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF058919; AAC13627.1; -.
DR   EMBL; AL161472; CAB80871.1; -.
DR   EMBL; AY059944; AAL24426.1; -.
DR   EMBL; BT000066; AAN15385.1; -.
DR   PIR; T01226; T01226.
DR   HSSP; P11586; 1A4I.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   360 AA;  38741 MW;  15D254E1DDD2F6B2 CRC64;

  Query Match             83.3%;  Score 1238;  DB 10;  Length 360;
  Best Local Similarity   79.1%;  Pred. No. 1.1e-95;
  Matches  231;  Conservative   36;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :|||| |||:||:|| :|:::||::|||:|||||||||| |||| ||||||||||::||
Db         69 AIVIDGKAVAKKIRDEITIEVSRMKESIGVIPGLAVILVGDRKDSATYVRNKKKACDSVG 128

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||:|| | |||||:|||| :: || |||||||||||||| ||:::|||||||||||||
Db        129 IKSFEVRLAEDSSEEEVLKSVSGFNDDPSVHGILVQLPLPSHMDEQNILNAVSIEKDVDG 188

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||||||:||:| |||||||||:|||||| :||||||||||||||||||||||||||
Db        189 FHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNIEIKGKRAVVIGRSNIVGMPAALLLQ 248

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              : :|||||:|| || ||||||:|||:|:|||  |:||||||||| :||||||:|| :
Db        249 REDATVSIIHSRTKNPEEITREADIIISAVGQPNMVRGSWIKPGAVLIDVGINPVEDPSA 308

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
              |||||||:|||||||:|  ||||||||||||||||||| |||||| |:
Db        309 ARGYRLVGDICYEEASKVASAITPVPGGVGPMTIAMLLSNTLTSAKRIHNFQ 360


RESULT 2
Q8LEA3
ID   Q8LEA3      PRELIMINARY;      PRT;   299 AA.
AC   Q8LEA3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   5,10-methylenetetrahydrofolate dehydrogenase:5,10-
DE   methenyltetrahydrofolate cyclohydrolase, putative.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 0:0-0(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RT   "Full-Length cDNA from Arabidopsis thaliana.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY085538; AAM62762.1; -.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
SQ   SEQUENCE   299 AA;  31566 MW;  C934C84E91D7BAF7 CRC64;

  Query Match             69.7%;  Score 1035;  DB 10;  Length 299;
  Best Local Similarity   67.4%;  Pred. No. 9.5e-79;
  Matches  194;  Conservative   45;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              ||||||| :|  || ||| |:  : :  | ||||||::||||||||||  |:|||  ||
Db          9 AKIIDGKAIAHTIRSEIAEEVRGLSEKHGKVPGLAVVIVGSRKDSQTYVNTKRKACAEVG 68

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||::| |||: || :::  :   ||:| |||||||||||| |:|:|||| |:||:|||||
Db         69 IKSFDVGLPEEVSEADLISKVHELNSNPDVHGILVQLPLPKHINEENILGAISIDKDVDG 128

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:||:| |:|||||||:||| | ||:|||:||||:|||||||:| :|||
Db        129 FHPLNIGKLAMKGREPLFLPCTPKGCLELLARSGVKIKGQRAVVVGRSNIVGLPVSLLLL 188

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:|||: |||:|| || | |:||||||| | |::::|:|||||||:|||| | | ||
Db        189 KADATVTTVHSHTKDPEAIIREADIVIAACGQAHMIKGNWIKPGAAVIDVGTNAVSDPSK 248

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
              |||||||  : ||||:|| |||||||||||:|||| ||:: |||:
Db        249 KSGYRLVGDVDFAEASKVAGFITPVPGGVGPMTVAMLLRNTVDGAKRV 296


RESULT 3
Q9ZTV0
ID   Q9ZTV0      PRELIMINARY;      PRT;   294 AA.
AC   Q9ZTV0;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   5,10-methylenetetrahydrofolate dehydrogenase-5,10-
DE   methenyltetrahydrofolate cyclohydrolase.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chen L., Nargang N.E., Cossins E.A.;
RT   "a cDNA sequence encoding pea leaf cytosolic bifunctional 5,10-
RT   methylenetetrahydrofolate dehydrogenase-5,10-methenyltetrahydrofolate
RT   cyclohydrolase.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
```

RA Chen L., Nargang F.E., Cossins E.A.;
RT "Isolation and sequencing of a cDNA encoding pea cytosolic
RT bifunctional 5,10-methylenetetrahydrofolate dehydrogenase:5,10-
RT methenyltetrahydrofolate cyclohydrolase. Comparisons with the yeast,
RT mammalian, and bacterial proteins.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF030516; AAD01907.1; -.
DR EMBL; AJ011589; CAB56756.1; -.
DR PIR; T50664; T50664.
DR HSSP; P11586; 1A4I.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THF_Dhg_Cyh.
DR Pfam; PF00763; THF_DHG_CYH; 1.
DR Pfam; PF02882; THF_DHG_CYH_C; 1.
DR PRINTS; PR00085; THFDHDRGNASE.
DR ProDom; PD002300; THF_Dhg_Cyh; 1.
DR PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW Hydrolase; Oxidoreductase.
SQ SEQUENCE 294 AA; 31344 MW; 5D5066E8BA880690 CRC64;

Query Match 69.6%; Score 1034; DB 10; Length 294;
Best Local Similarity 67.2%; Pred. No. 1.1e-78;
Matches 193; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              | :|||| ||: || ||| |:  :    | |||||::||:|||||:||  |:|||  :|
Db          2 ATVIDGKAVAQTIRSEIADEVRLLSQKYGKVPGLAVVIVGNRKDSQSYVGMKRKACAELG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
              |||:::|||||:|| |::|::   |::| |||||||||| |:|:| :|  :|| |||||
Db         62 IKSFDIDLPEDASEAEIIKNVHELNANPDVHGILVQLPLPKHVNEEKVLTEISISKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
              |||||||:|||:|||| |:||||| |:||| | || ||||:|||:|||||||:||:|||
Db        122 FHPLNIGKLAMKGRDPLFLPCTPKACLELLSRSGVSIKGKKAVVVGRSNIVGLPASLLLL 181

Qy        181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
              ||:|||:||||:| :|| | |:||||||| | | :::|||||||||:|||| | ||||
Db        182 KADATVTIVHSHTSQPETIIREADIVIAAAGQAKMIKGSWIKPGAAVIDVGTNSVDDPTR 241

Qy        241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
              |||||||| :|||||:|| |||||||||||||: ||| |||| |||
Db        242 KSGYRLVGDVDFEEEASKVAGWITPVPGGVGPMTVTMLLKNTLEGAKR 288
```

RESULT 4
Q9LHH7
ID Q9LHH7 PRELIMINARY; PRT; 299 AA.
AC Q9LHH7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 5,10-methylenetetrahydrofolate
DE dehydrogenase/5,10-methenyltetrahydrofolate cyclohydrolase
DE (5,10-methylenetetrahydrofolate
DE dehydrogenase:5,10-methenyltetrahydrofolate cyclohydrolase, putative,
DE 44272-46007) (At3g12290).
GN F28J15.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:820-822(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP002047; BAB03138.1; -.
DR EMBL; AC069472; AAG51064.1; -.
DR EMBL; BT004612; AAO42858.1; -.
DR HSSP; P11586; 1A4I.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THF_Dhg_Cyh.
DR Pfam; PF00763; THF_DHG_CYH; 1.
DR Pfam; PF02882; THF_DHG_CYH_C; 1.
DR PRINTS; PR00085; THFDHDRGNASE.
DR ProDom; PD002300; THF_Dhg_Cyh; 1.
DR PROSITE; PS00767; THF_DHG_CYH_2; 1.
SQ SEQUENCE 299 AA; 31589 MW; CA92CBEB3271BAF1 CRC64;

Query Match 69.3%; Score 1030; DB 10; Length 299;
Best Local Similarity 67.0%; Pred. No. 2.5e-78;
Matches 193; Conservative 46; Mismatches 49; Indels 0; Gaps 0;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              ||||||| :|  || ||| |:  : :  | ||||||::||||||||||||  |:|||  ||
Db          9 AKIIDGKAIAHTIRSEIAEEVRGLSEKHGKVPGLAVVIVGSRKDSQTYVNTKRKACAEVG 68
```

```
Qy          61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
               |||::| |||: || ::: :  ||:| |||||||||| |:|:|:|| |:||:|||||
Db          69 IKSFDVGLPEEVSEADLISKVHELNSNPDVHGILVQLPLPKHINEEHILGAISIDKDVDG 128

Qy         121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
               |||||||:|||:||:| |:|||||||:||| | ||:|||:||||:|||||||:| :|||
Db         129 FHPLNIGKLAMKGREPLFLPCTPKGCLELLARSGVKIKGQRAVVVGRSNIVGLPVSLLLL 188

Qy         181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
               ||:|||: |||:|| || | |:||||||| | |::::|:|||||||:|||| | | ||
Db         189 KADATVTTVHSHTKDPEAIIREADIVIAACGQAHMIKGNWIKPGAAVIDVGTNAVSDPSK 248

Qy         241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
                  |||||||| : ||||:|| |||||||||||||:|||| ||:: |||:
Db         249 KSGYRLVGDVDFAEASKVAGFITPVPGGVGPMTVAMLLRNTVDGAKRV 296


RESULT 5
Q9ZVI8
ID   Q9ZVI8      PRELIMINARY;      PRT;   350 AA.
AC   Q9ZVI8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Methylenetetrahydrofolate dehydrogenase.
GN   AT2G38660.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005499; AAC67352.1; -.
DR   PIR; G84807; G84807.
DR   HSSP; P11586; 1A4I.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
SQ   SEQUENCE   350 AA;  37804 MW;  51F14446F907C1A3 CRC64;

  Query Match             68.4%;  Score 1017;  DB 10;  Length 350;
  Best Local Similarity   67.5%;  Pred. No. 3.9e-77;
  Matches  193;  Conservative   36;  Mismatches   57;  Indels    0;  Gaps    0;

Qy           3 IIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGIK 62
               :||| ::|::|| :|  |: ||| |:| ||||||:||| ::|||||||| ||||  |||
Db          63 VIDGNVIAEEIRTKIISEVGKMKKAVGKVPGLAVVLVGEQRDSQTYVRNKIKACEETGIK 122

Qy          63 SYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGFH 122
               |    |||| :| ::: :  || | |:||||||||||| |:|:  ||| | :||||||||
Db         123 SVLAELPEDCTEGQIISVLRKFNEDTSIHGILVQLPLPQHLNESKILNMVRLEKDVDGFH 182

Qy         123 PLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQKA 182
               |||:| |||:||:| || ||||||:||| | |||| || ||||||||||:| :||||:
Db         183 PLNVGNLAMRGREPLFVSCTPKGCVELLIRTGVEIAGKNAVVIGRSNIVGLPMSLLLQRH 242

Qy         183 NATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESPR 242
               :|||| ||: || || |||:||||||| |: |||||||:|||| :|||| ||:|
Db         243 DATVSTVHAFTKDPEHITRKADIVIAAAGIPNLVRGSWLKPGAVVIDVGTTPVEDSSCEF 302

Qy         243 GYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
               |||||||||||||  :|  |||||||||||||| ||| ||||:||||
Db         303 GYRLVGDVCYEEALGVASAITPVPGGVGPMTITMLLCNTLEAAKRI 348


RESULT 6
O65269
ID   O65269      PRELIMINARY;      PRT;   310 AA.
AC   O65269;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   F6N23.28 protein (Putative tetrahydrofolate synthase).
GN   F6N23.28 OR AT4G00600.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   WASHU;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Geisel C.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Waterston R.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF058919; AAC13632.1; -.
DR   EMBL; AL161472; CAB80869.1; -.
DR   PIR; T01224; T01224.
DR   HSSP; P11586; 1A4I.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
SQ   SEQUENCE   310 AA;  33696 MW;  9A71117B64896FBC CRC64;

  Query Match             67.9%;  Score 1009;  DB 10;  Length 310;
  Best Local Similarity   66.4%;  Pred. No. 1.5e-76;
  Matches  194;  Conservative   33;  Mismatches   29;  Indels   36;  Gaps    1;

Qy           1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
               | :|||  ||:||::| :|:::||::|||||
Db          55 AIVIDGKAEAKKIRDDIKIEVSRMKESIGVVPA---------------------------- 87
```

```
Qy          61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
                ||||||:||||::: || |||||||:||||||| ||:::|||||||||||||||
Db          88 ---------EDSSEEEVLKYVSGFNDDPSVHGVLVQLPLPSHMDEQNILNAVSIEKDVDG 138

Qy         121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
               |||||||||||:||:| |||||||||:|||||| :| ||||||||||||||||||||||
Db         139 FHPLNIGRLAMRGREPLFVPCTPKGCIELLHRYNIEFKGKRAVVIGRSNIVGMPAALLLQ 198

Qy         181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
               | :|||||:|| |  |||:||||||:|:|||  |:|||||||||| :||||| ||:|| :
Db         199 KEDATVSIIHSRTMNPEELTRQADILISAVGKPNMVRGSWIKPGAVLIDVGIKPVEDPSA 258

Qy         241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHKFK 292
                 | |||||:|| ||||||  |||||| ||||||||||||||| ||||||  |:
Db         259 AGGERLVGDICYVEASKIASAITPVPGDVGPMTIAMLLSNTLTSAKRIHNFQ 310
```

RESULT 7
Q9FRB2
ID Q9FRB2 PRELIMINARY; PRT; 291 AA.
AC Q9FRB2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to Pisum sativum methylenetetrahydrofolate dehydrogenase
DE (NADP+) (EC 1.5.1.5) (AF030516).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Chow T.-Y., Hsing Y.-I.C., Chen H.-H., Wu H.-P., Chao Y.-T.,
RA Liu S.-M., Hsiao Y.-Y., Huang J.-J., Lee P.-F., Su C.-L., Chen C.-S.,
RA Shaw J.-F.;
RT "Oryza sativa PAC P0001A07 genomics sequence.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC084218; AAG48834.1; -.
DR HSSP; P11586; 1A4I.
DR Gramene; Q9FRB2; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004488; F:methylenetetrahydrofolate dehydrogenase (NA. . .; IEA.
DR GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THF_Dhg_Cyh.
DR Pfam; PF00763; THF_DHG_CYH; 1.
DR Pfam; PF02882; THF_DHG_CYH_C; 1.
DR PRINTS; PR00085; THFDHDRGNASE.
DR ProDom; PD002300; THF_Dhg_Cyh; 1.
DR PROSITE; PS00767; THF_DHG_CYH_2; 1.
SQ SEQUENCE 291 AA; 30720 MW; 81591D0662638C9E CRC64;

Query Match 66.8%; Score 992.5; DB 10; Length 291;
Best Local Similarity 66.6%; Pred. No. 3.4e-75;
Matches 191; Conservative 39; Mismatches 56; Indels 1; Gaps 1;

```
Qy           1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
               |:||||| ||  || |:| ::| :  |  :||||||::||||||||||: |:|||  ||
Db           2 AQIIDGKAVAADIRREVAADVAALSSAHNLVPGLAVVIVGSRKDSQTYVQMKRKACAEVG 61

Qy          61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
               |:| :|:| || ||  ::  :   |:||:|||||||||||| |:|:| ||| :|:||||||
Db          62 IRSVDVDLAEDISEAALVAEVHRLNADPAVHGILVQLPLPKHINEEKILNEISLEKDVDG 121

Qy         121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQ 180
               |||||||:|||:|||| |:||||||||||| | || | ||||||:|||||||:| :|||
Db         122 FHPLNIGKLAMKGRDPLFLPCTPKGCMELLTRSGVTINGKRAVVVGRSNIVGLPVSLLLL 181

Qy         181 KANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPES 240
               ||:|||||||| |  || | |:||||||| | | :|   |||||||:|||| | : ||
Db         182 KADATVSIVHSRTPNPESIVREADIVIAAAGQAMMVH-YWIKPGAAVIDVGTNSISDPTR 240

Qy         241 PRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
               |||||||| : | ||:|| :|||||||||||:|||| ||:: |||
Db         241 KSGYRLVGDVDFAEVSKVAGHLTPVPGGVGPMTVAMLLKNTVDGAKR 287
```

RESULT 8
Q8W527
ID Q8W527 PRELIMINARY; PRT; 160 AA.
AC Q8W527;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Methylenetetrahydrofolate dehydrogenase/methylenetetrahydrofolate
DE cylcohydrolase isoform 2 (EC 3.5.4.9) (Fragment).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Bohnert H.;
RT "Plant one-carbon genes.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF439725; AAL33591.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004477; F:methenyltetrahydrofolate cyclohydrolase act. . .; IEA.
DR GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THF_Dhg_Cyh.
DR Pfam; PF02882; THF_DHG_CYH_C; 1.
DR PRINTS; PR00085; THFDHDRGNASE.
DR ProDom; PD002300; THF_Dhg_Cyh; 1.
DR PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW Hydrolase.
FT NON_TER 1 1
FT NON_TER 160 160
SQ SEQUENCE 160 AA; 16935 MW; 325D0A962AD26C2A CRC64;

Query Match 53.3%; Score 792; DB 10; Length 160;
Best Local Similarity 94.4%; Pred. No. 1.1e-58;
Matches 151; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```
Qy         115 EKDVDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMP 174
               |||||||||||||||||||||||||||||||||||||||:||||||||||||||||||| |
Db           1 EKDVDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRHGVEIKGKRAVVIGRSNIVGTP 60

Qy         175 AALLLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINP 234
               |||||||||||||||| || ||:|||||||||||||||||| |||||| ||||||||
Db          61 AALLLQKANATVSIVHSQTKNPEQITRQADIVIAAVGVANLVRGDWIKPGATIIDVGINP 120

Qy         235 VDDPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTI 274
               |||||||||||||||||||||||:|| ||||||||||||
Db         121 VDDPESPRGYRLVGDVCYEEASKVAGAITPVPGGVGPMTI 160
```

RESULT 9
Q9HFV0
ID Q9HFV0 PRELIMINARY; PRT; 937 AA.
AC Q9HFV0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE C1-THFS protein.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RA Langridge J., Bordoli R., Rachubinski R., Wozniak R., Smith J.,
RA Skilling J.;
RT "Direct software interpretation of de novo sequence from ESI-MS/MS
RT data for novel protein identification using BLAST.";

```
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF291429; AAG11417.1; -.
DR   HSSP; P11586; 1A4I.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0004329; F:formate-tetrahydrofolate ligase activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000559; Fmtethyd_synth.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF01268; FTHFS; 1.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00721; FTHFS_1; 1.
DR   PROSITE; PS00722; FTHFS_2; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
SQ   SEQUENCE   937 AA;  100193 MW;  DA3B0A16C9704D51 CRC64;

  Query Match             51.5%;  Score 765.5;  DB 3;  Length 937;
  Best Local Similarity   54.2%;  Pred. No. 2.2e-55;
  Matches  156;  Conservative   39;  Mismatches   92;  Indels    1;  Gaps    1;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDA-IGVVPGLAVILVGSRKDSQTYVRNKKKACEAV 59
              | ||||  :|| ||| :  ||| ::       | | :| || | || |||| | || :
Db          2 ASIIDGNSIAKNIRERLHTEIADVQQTNPRFQPSLVIIQVGDRPDSSTYVRMKLKAADEA 61

Qy         60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVD 119
               ||      | |: |: |:|: |   | |||| |:|||||| |:::  | |||: :||||
Db         62 AIKCELKKLDENVSQHELLREITHLNEDPSVDGVLVQLPLPKHIDETTITNAVAADKDVD 121

Qy        120 GFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLL 179
              || ||||| || :| :| || ||||| ||||    |::: || |||:|||:||| | : ||
Db        122 GFGPLNIGELAKRGGNPSFVACTPKGVMELLKVAGIDVAGKTAVVLGRSDIVGAPVSYLL 181

Qy        180 QKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPE 239
              : |:|||::||| |    || : ||||:||:|    |:| |:|||| :|||| | : |
Db        182 RNADATVTVVHSKTANIPEIVKTADIVVAAIGQPGFVKGEWLKPGAVVIDVGTNYIPDDT 241

Qy        240 SPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                | |||||| :| |  :|  ||||||||||||:|||| | : ||||
Db        242 KKSGQRLVGDVDFESAKNVASHITPVPGGVGPMTVAMLLDNVVLSAKR 289


RESULT 10
Q92BZ4
ID   Q92BZ4      PRELIMINARY;      PRT;   284 AA.
AC   Q92BZ4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FolD protein.
GN   FOLD OR LIN1397.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / Serovar 6a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596168; CAC96628.1; -.
DR   PIR; AD1607; AD1607.
DR   ListiList; LIN01397; -.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Complete proteome.
SQ   SEQUENCE   284 AA;  30863 MW;  6777F2609936988B4 CRC64;

  Query Match             51.4%;  Score 763.5;  DB 16;  Length 284;
  Best Local Similarity   52.2%;  Pred. No. 5.9e-56;
  Matches  151;  Conservative   59;  Mismatches   70;  Indels    9;  Gaps    4;

Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              :||||| :||:|:|::  |:|::    |  |||||:|||  : |:|||||:|  |  |:
Db          3 EIIDGKKLAKEIQEKVTSEVAELVKQ-GKKPGLAVVLVGDNQASRTYVRNKQKRTEEAGM 61

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
               ||  : |||  :|:::|| :   | | ::||||||||| |:::| :::|:| :||||||
Db         62 KSVLIELPETVTEEKLLKVVEELNEDNTIHGILVQLPLPKHISEEKVIDAISFDKDVDGF 121

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              ||:|:| | : |:|  |||||| | :||:   | :|:|||||||||||||| | | ||
Db        122 HPVNVGNLFI-GKDS-FVPCTPAGIIELIKSTGTQIEGKRAVVIGRSNIVGKPVAQLLLN 179

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              ||||:| || ||   :: ::|||:: | |:|  |:  :|||||  :||||:      :
Db        180 ENATVTIAHSRTKDLPQVAKEADILVVATGLAKFVKKEYIKPGAIVIDVGM------DRD 233

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHK 290
              :| ||| :::  : || |||||||||||| |||:|||::|||| |
Db        234 ENNKLCGDVDFDDVKEQAGFITPVPGGVGPMTITMLLANTLKAAKRIWK 282


RESULT 11
Q8Y7C5
ID   Q8Y7C5      PRELIMINARY;      PRT;   284 AA.
AC   Q8Y7C5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FolD protein.
GN   FOLD OR LMO1360.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-e / Serovar 1/2a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL591978; CAC99438.1; -.
DR   PIR; AH1244; AH1244.
DR   ListiList; LMO01360; -.
```

```
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Complete proteome.
SQ   SEQUENCE   284 AA;  30907 MW;  0413362CF693598B CRC64;

  Query Match             50.8%;  Score 755.5;  DB 16;  Length 284;
  Best Local Similarity   51.6%;  Pred. No. 2.8e-55;
  Matches  149;  Conservative   60;  Mismatches   71;  Indels    9;  Gaps    4;

Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              :|||||  :||:|:|::  |:|::    |  |||||:|||  : |:|||||:|  |  |:
Db          3 EIIDGKKLAKEIQEKVTREVAELVKE-GKQPGLAVVLVGDNQASRTYVRNKQKRTEEAGM 61

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
              ||  : |||: :|:::|  :   | | ::|||||||||| |:::| ::: :| :||||||
Db         62 KSVLIELPENVTEEKLLSVVEELNEDKTIHGILVQLPLPEHISEEKVIDTISFDKDVDGF 121

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              ||:|:|  | : |:|  ||||||  | :||:   | :|:|||||||||||||| | | ||
Db        122 HPVNVGNLFI-GKDS-FVPCTPAGIIELIKSTGTQIEGKRAVVIGRSNIVGKPVAQLLLN 179

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              ||||:| || ||   :: ::|||:: | |:|  |:  :|||||  :||||:          :
Db        180 ENATVTIAHSRTKDLPQVAKEADILVVATGLAKFVKKDYIKPGAIVIDVGM------DRD 233

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHK 290
                :| ||| :::  : || |||||||||||| |||:|||::|||| |
Db        234 ENNKLCGDVDFDDVVQEAGFITPVPGGVGPMTITMLLANTLKAAKRIWK 282


RESULT 12
Q8WZJ7
ID   Q8WZJ7      PRELIMINARY;      PRT;   937 AA.
AC   Q8WZJ7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   C-1-tetrahydrofolate synthase.
GN   SPBC839.16.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972h-;
RA   Lyne M., Purnelle B., Goffeau A., Wood V., Rajandream M.A.,
RA   Barrell B.G., Saunders D., Harris D.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL096796; CAB46709.1; -.
DR   PIR; T40723; T40723.
DR   GeneDB_SPombe; SPBC839.16; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0004329; F:formate-tetrahydrofolate ligase activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000559; Fmtethyd_synth.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF01268; FTHFS; 1.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00721; FTHFS_1; 1.
SQ   SEQUENCE   937 AA;  101202 MW;  70FF8700FD023C90 CRC64;

  Query Match             50.4%;  Score 749.5;  DB 3;  Length 937;
  Best Local Similarity   50.0%;  Pred. No. 4.8e-54;
  Matches  147;  Conservative   50;  Mismatches   90;  Indels    7;  Gaps    2;

Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVP----GLAVILVGSRKDSQTYVRNKKKAC 56
              | :::|  :|:::|||:  :|: :|    | |      | :| || |:|| ||| | :|
Db          2 ALLLEGTSLARKVREELREQISSIK---SVDPYFNVSLKIIQVGGREDSNVYVRMKTRAA 58

Qy         57 EAVGIKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEK 116
                ||    || ||| :| ::|  |  || ||:||||:|||||| |:|:: |  ||: ||
Db         59 NEAGISCEHVNFPEDITEYDLLLAIKGFNEDPTVHGIIVQLPLPAHINEQIITEAVAPEK 118

Qy        117 DVDGFHPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAA 176
              |||||   |:|:|  :   | |  ||||| | :|  ||: ::|| ||||||||||| | :
Db        119 DVDGFCETNLGKLTKREGQPLFTACTPKGIMCILKHYGINVQGKHAVVIGRSNIVGRPMS 178

Qy        177 LLLQKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVD 236
              :||:|||||:: || |:   :| | ||||:||:|: : |:  |:| |   ||||| :
Db        179 ILLEKANATVTLCHSKTESIADIVRTADIVVAAIGIPHFVKADWLKKGVVAIDVGINSIP 238

Qy        237 DPESPRGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRIHK 290
              |     |||| ||: :| | ::|  |||||| ||||:|||| | :||| |  |
Db        239 DATKKSGYRLTGDIDFENAKEVASAITPVPGSVGPMTVAMLLQNVVESAVRFRK 292


RESULT 13
Q99YX2
ID   Q99YX2      PRELIMINARY;      PRT;   284 AA.
AC   Q99YX2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative bifunctional methylenetetrahydrofolate dehydrogenase /
DE   methenyltetrahydrofolate cyclohydrolase (EC 1.5.1.5).
GN   FOLD OR SPY1502 OR SPYM18_1520.
OS   Streptococcus pyogenes, and
OS   Streptococcus pyogenes (serotype M18).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1314, 186103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX   MEDLINE=21192684; PubMed=11296296;
RA   Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA   Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA   Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J.,
RA   Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT   "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MGAS8232 / Serotype M18;
RX   MEDLINE=21927593; PubMed=11917108;
RA   Smoot J.C., Barbian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA   Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA   Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA   Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT   "Genome sequence and comparative microarray analysis of serotype M18
RT   group A Streptococcus strains associated with acute rheumatic fever
RT   outbreaks.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
DR   EMBL; AE006583; AAK34300.1; -.
DR   EMBL; AE010067; AAL98090.1; -.
DR   HSSP; P11586; 1A4I.
DR   GO; GO:0004488; F:methylenetetrahydrofolate dehydrogenase (NA. . .; IEA.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
```

```
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Hypothetical protein; Oxidoreductase; Complete proteome.
SQ   SEQUENCE   284 AA;  31082 MW;  E81EA697A4CBBE35 CRC64;

  Query Match           50.0%;  Score 743.5;  DB 16;  Length 284;
  Best Local Similarity   51.4%;  Pred. No. 2.9e-54;
  Matches  149;  Conservative   55;  Mismatches   77;  Indels    9;  Gaps    4;

Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              ::|||| :|:::::|:| ::  :|   |:||||||||||    || |||||::|  ||
Db          3 ELIDGKALAQKMQQELAAKVNNLKQKKGIVPGLAVILVGDDPASQVYVRNKERAALTVGF 62

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
              ||  | | |   ::|::  |  :|:| ::||||||||||:|:||: |: |:  :||||||
Db         63 KSETVRLSEFICQEELIAVIERYNADNTIHGILVQLPLPNHINDKKIILAIDPKKDVDGF 122

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              ||:| | |   || |  ||||| | ||||  | | ::|| ||:||||||| | | ||
Db        123 HPMNTGHL-WSGR-PLMVPCTPSGIMELLREYNVNLEGKHAVIIGRSNIVGKPMAQLLLD 180

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              ||||:: || |:: ||: | ||::| |:|  : :   :|| || :||||:|  |:
Db        181 KNATVTLTHSRTRQLEEVCRCADVLIVAIGQGHFITKQYIKDGAIVIDVGMNRDDNG--- 237

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR-IHK 290
                :|:||| ::| :::|  ||||||||||||||||  | :|| |  ||
Db        238 ---KLIGDVAFDEVAEVAAKITPVPGGVGPMTIAMLLEQTYQSALRSTHK 284


RESULT 14
Q8K6S8
ID   Q8K6S8      PRELIMINARY;      PRT;   284 AA.
AC   Q8K6S8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative methylenetetrahydrofolate dehydrogenase.
GN   FOLD OR SPYM3_1157 OR SPS0705.
OS   Streptococcus pyogenes (serotype M3).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=198466;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MGAS315 / Serotype M3;
RX   MEDLINE=22133808; PubMed=12122206;
RA   Beres S.B., Sylva G.L., Barbian K.D., Lei B., Hoff J.S.,
RA   Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA   Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA   Schlievert P.M., Musser J.M.;
RT   "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT   phage-encoded toxins, the high-virulence phenotype, and clone
RT   emergence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SSI-1 / Serotype M3;
RA   Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,
RA   Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,
RA   Hayashi H., Hamada S.;
RT   "The genome of invasive Streptococcus pyogenes; a comparative analysis
RT   of S. pyogenes SSI-1, SF370 and MGAS8232.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE014159; AAM79764.1; -.
DR   EMBL; AP005143; BAC63800.1; -.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
```

```
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   284 AA;  31096 MW;  E81F1727A4CBBF34 CRC64;

  Query Match           50.0%;  Score 743.5;  DB 16;  Length 284;
  Best Local Similarity   51.4%;  Pred. No. 2.9e-54;
  Matches  149;  Conservative   55;  Mismatches   77;  Indels    9;  Gaps    4;

Qy          2 KIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61
              ::|||| :|:::::|:| ::  :|   |:||||||||||    || |||||::|  ||
Db          3 ELIDGKALAQKMQQELAAKVNNLKQKKGIVPGLAVILVGDDPASQVYVRNKERAALTVGF 62

Qy         62 KSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDGF 121
              ||  | | |   ::|::  |  :|:| ::||||||||||:|:||: |: |:  :||||||
Db         63 KSETVRLSEFICQEELIAVIERYNADNTIHGILVQLPLPNHINDKKIILAIDPKKDVDGF 122

Qy        122 HPLNIGRLAMQGRDPFFVPCTPKGCMELLHRYGVEIKGKRAVVIGRSNIVGMPAALLLQK 181
              ||:| | |   || |  ||||| | ||||  | | ::|| ||:||||||| | | ||
Db        123 HPMNTGHL-WSGR-PLMVPCTPSGIMELLREYNVNLEGKHAVIIGRSNIVGKPMAQLLLD 180

Qy        182 ANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPESP 241
              ||||:: || |:: ||: | ||::| |:|  : :   :|| || :||||:|  |:
Db        181 KNATVTLTHSRTRQLEEVCRCADVLIVAIGQGHFITKQYIKEGAIVIDVGMNRDDNG--- 237

Qy        242 RGYRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR-IHK 290
                :|:||| ::| :::|  ||||||||||||||||  | :|| |  ||
Db        238 ---KLIGDVAFDEVAEVAAKITPVPGGVGPMTIAMLLEQTYQSALRSTHK 284


RESULT 15
Q8FVP6
ID   Q8FVP6      PRELIMINARY;      PRT;   299 AA.
AC   Q8FVP6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FolD bifunctional protein.
GN   FOLD OR BRA0781.
OS   Brucella suis.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29461;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1330 / Biovar 1;
RX   MEDLINE=22247741; PubMed=12271122;
RA   Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA   Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA   Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA   Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA   Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA   Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT   "The Brucella suis genome reveals fundamental similarities between
RT   animal and plant pathogens and symbionts.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR   EMBL; AE014572; AAN33961.1; -.
DR   TIGR; BRA0781; -.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR   InterPro; IPR000672; THF_Dhg_Cyh.
DR   Pfam; PF00763; THF_DHG_CYH; 1.
DR   Pfam; PF02882; THF_DHG_CYH_C; 1.
DR   PRINTS; PR00085; THFDHDRGNASE.
DR   ProDom; PD002300; THF_Dhg_Cyh; 1.
DR   PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR   PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW   Complete proteome.
SQ   SEQUENCE   299 AA;  31231 MW;  7446B7B964FB967A CRC64;

  Query Match           49.9%;  Score 741;  DB 16;  Length 299;
  Best Local Similarity   51.6%;  Pred. No. 5e-54;
```

Matches 149; Conservative 49; Mismatches 89; Indels 2; Gaps 2;

```
Qy          1 AKIIDGKLVAKQIREEIAVEIAKMKDAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVG 60
              |::|||| :|: :   :  |  |:  | |||||:||::||    || || :| :  :  |
Db          2 AQLIDGKKLAEDVVSTVKTETEKLVAATGVVPGIAVVIVGEDPASQVYVASKSRKAKECG 61

Qy         61 IKSYEVNLPEDSSEDEVLKHIATFNSDPSVHGILVQLPLPHHMNDENILNAVSIEKDVDG 120
               | : :||| :|| |:|  |   |:||::|||||||||| |::   ::  :: ||||||
Db         62 FHSVQHDLPETASEQELLNLIEGLNNDPAIHGILVQLPLPGHIDSGRVIQTIAPEKDVDG 121

Qy        121 FHPLNIGRLAMQGRDPFFVPCTPKGCMELLHR-YGVEIKGKRAVVIGRSNIVGMPAALLL 179
              || :|:|:|    :  |||||| | | :: | :| :: |  |||||||||| |   ||
Db        122 FHFINVGKLGTGEVETAFVPCTPAGAMIMIERVHGRDLSGLNAVVIGRSNIVGKPMFNLL 181

Qy        180 QKANATVSIVHSNTKKPEEITRQADIVIAAVGVANLVRGSWIKPGAAIIDVGINPVDDPE 239
              |||||:: || ||    | | |||::||||   :|:| |:|||| :|||||  :  ||
Db        182 LAANATVTVAHSRTKDLPAIARNADILVAAVGRPQMVKGDWVKPGATVIDVGINRIPAPE 241

Qy        240 SPRG-YRLVGDVCYEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKR 287
                |  |||||| : || |:|| |||||||||||||||||::||| :| |
Db        242 RGEGKTRLVGDVDFAEAEKVAGAITPVPGGVGPMTIAMLMANTLTAACR 290
```

Search completed: July 24, 2004, 06:38:59
Job time : 87 secs